A multi-country approach for attributing human salmonellosis to animal reservoirs

Global perspectives and application of surveillance data from the European Union

de Knegt, Leonardo; Hald, Tine; Pires, Sara Monteiro

Publication date:
2013

Document Version
Publisher's PDF, also known as Version of record

A multi-country approach for attributing human salmonellosis to animal reservoirs: Global perspectives and application of surveillance data from the European Union

Leonardo Victor de Knegt
PhD Thesis
January 2013

DTU Food
National Food Institute

A multi-country approach for attributing human salmonellosis to animal reservoirs: Global perspectives and application of surveillance data from the European Union

Ph.D. Thesis

Leonardo Víctor de Knegt

January 2013

National Food Institute, Technical University of Denmark

Supervisors:

Tine Hald, DVM, PhD, Research group leader
Epidemiology and Risk Modelling Group, Division of Epidemiology and Microbial Genomics
National Food Institute, Technical University of Denmark, Søborg, Denmark

Sara Monteiro Pires, DVM, PhD, Researcher
Epidemiology and Risk Modelling Group, Division of Epidemiology and Microbial Genomics
National Food Institute, Technical University of Denmark, Søborg, Denmark

Assessment committee:

Rene S. Hendriksen, PhD, Senior researcher
Antimicrobial Resistance Group, Division of Epidemiology and Microbial Genomics
National Food Institute, Technical University of Denmark, Lyngby, Denmark

Dörte Dorothea Vera Döpfer, DVM, PhD, Assistant Professor
Food Animal Medicine Section of the School of Veterinary Medicine
University of Wisconsin, Madison, United States of America

Antonia Ricci, DVM, Head of Laboratory
National reference laboratory for *Salmonella*
OIE Reference laboratory for salmonellosis
Food safety department
Epidemiology laboratory
Laboratory of Public Health and Risk Analysis
Istituto Zooprofilattico Sperimentale delle Venezie, Venice, Italy

PhD scholarship granted with one third of the financing from the National Food Institute and two thirds from the Technical University of Denmark.

ISBN: 978-87-92763-46-4
Cover photo: Colourbox
Cover art: Leonardo de Knegt and Susanne Carlsson
Printed by Rosendahl Schultz Grafisk A/S

Preface

The work presented in this thesis was conducted between April 2009 and June 2012 at the National Food Institute, Technical University of Denmark.

The idea behind this PhD was to contribute to the path being built towards source attribution of foodborne salmonellosis at a Global level, and it was inspired by the activities of the World Health Organization Foodborne Disease Epidemiology Reference Group (FERG) and of the Global Foodborne Infections Network (GFN). Due to the current data availability when starting the project, the European Union was chosen as the best scenario to rehearse multi-country approaches that can later be adapted to different realities.

The European Union model was developed as part of a contract CT/EFSA/Zoonoses/2010/02 between the European Food Safety Authority and the DTU National Food Institute, in relation to EFSA Question n° EFSA-Q-2010-00685.

The comparison between the Danish and EU models was performed as per request of the Danish Food Administration (Fødevarestyrelsen).

Oticon Fonden granted me partial financial support to attend to the USDA Food Safety Education Conference and the Second Formal Meeting of the FERG Country Studies Task Force in Atlanta, March 2010, as well as to my external stay at the WHO Department of Food Safety and Zoonoses (WHO/FOS) in Geneva, from July to December 2010.

The author hopes that the materials and methods presented in this thesis are useful to improve the quality of life of populations threatened by infections of foodborne transmission.

Søborg, January 2013
Leonardo de Knegt

Acknowledgements

This thesis is dedicated to Ana, my best friend and companion for life, and to Laura, double-thesis baby and love of our lives.

I would like to thank my parents for the unconditional support and for learning to live so far away from a branch of their family.

I also thank Tine for welcoming me into the group, for the supervision, teaching, motivation, discussions and exchange of Star Wars references; Sara for introducing me to the attribution model and WinBUGS; Antônio for ultimately bringing me to DTU and introducing me to SAS; Henrik Wegener, for welcoming us and helping us set up our home in a strange land; my PhD student colleagues from this and other departments or schools, as well as colleagues from Epi Modelling, Zoonosescentret and La Résistance for the help, discussions and fun; Håkan and Maarten for the informal supervision, helpful tips, hopjes and stroopwaffels; Anne Kjærgaard for being my rescuer in all human resources matters; Claudia and the WHO/FOS group for welcoming me and making me feel part of the group, and particularly to Tim, Linda, Stefania and Tanja for the true friendship; the Brazilian FETP (Episus), specially Beth, Wildo and Douglas, for turning me into a real epidemiologist and supporting my decision of coming to Denmark; and the Reverend Thomas Bayes and the "Mr. Price" who actually published his essay, starting this whole mess.

I acknowledge Frank Boelart and Giusi Amore for the co-work on the model report and the help with the data, and Timour Koupeev for the help with the EUROSTAT data.

The European Food Safety Authority and the European Centre for Disease Control and Prevention are acknowledged for providing the *Salmonella* data for the EU model.

```
/* Final thanks */

data final_thanks; set egtask.friendsfamilyandcolleagues;
if you = 'Those around me that helped or supported in any sense'
then message = 'Thank you for the friendship, support and for how much we grew
together during these three years';
place = 'Søborg, Denmark';
month = 'January';
year = '2013';
run;

/* End of acknowledgements */
```

Summary

This thesis presents a mathematical modeling approach to estimate the contribution of four animal reservoirs of the food chain to the occurrence of salmonellosis cases in humans in the European Union (Part I). In addition, an alternative and more explorative approach based on expert elicitation is attempted in order to extrapolate results to countries with less data availability, as a first step to perform source attribution of *Salmonella* in a more global perspective (Part II).

Cases of salmonellosis in humans were attributed to travel, foodborne outbreaks and four food animal reservoirs, namely pigs, broilers, turkeys and laying hens, using a Bayesian model based on microbial subtyping in 24 countries of the European Union. The chosen approach is recognized as data intensive, requiring numbers for *Salmonella* occurrence in food-producing animals, reported human cases, information on possibility of infection abroad (from here on referred to as "travel information"), human cases originating from outbreaks with and without a confirmed source and amounts of the meat or eggs available for consumption in each country. Thus, thorough data management, analysis and validation were required to produce a dataset containing standardized information for all countries (Manuscript I).

Data on reported human cases were provided by the European Centre for Disease Prevention and Control (ECDC) through the European Food Safety Authority (EFSA). *Salmonella* prevalences in animals were obtained from the EU-wide baseline studies (BS) (pigs and turkeys) conducted by EFSA and from the results of the harmonized monitoring (broiler and laying hens) reported in the European Union Summary Report (EUSR), as published by EFSA. Information on outbreaks was also provided by EFSA. The amount of food available for consumption was calculated based on trade data obtained from the European Statistical Office (EUROSTAT) and complemented with information from the Association of Poultry Processors and Poultry Trade in the European Union Countries (AVEC). Common limitations included non-participation in all BS, non-reporting of outbreaks or travel information, non-reporting of serovar-specific information, non-reporting of case-based data and non-availability of trade data on EUROSTAT. In order to standardize the information available, cases without travel information were assumed to be domestic; cases without specific serovar information were redistributed according to serovar proportions observed in the same dataset or other reference documents; missing trade information was estimated based on previous years, and non-participation in a BS was supplied, where possible, with data from the EUSR. When the lack of original data was considered too extreme to the point of compromising the attribution results, countries were excluded. The resulting dataset comprised Austria, Belgium, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, the Netherlands, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden and the United Kingdom. Three countries were included in the initial analysis, but were excluded from the final dataset. Those were: Bulgaria, which presented 100% of human cases without serovar detailing; and Romania, which only participated in one BS and had not enough surrogate data to be retrieved from the EUSR, besides reporting a large parcel of cases without serovar information (Manuscript I).

A Bayesian modeling approach which compares the occurrence of serovars in humans with the occurrence of the same serovars in animals of the food-chain was used to estimate the contribution of each of these reservoirs, travel and outbreaks to the number of human cases of salmonellosis in the 24 countries present in the dataset mentioned above (Manuscript II). Laying hens (i.e. eggs) were estimated to be the most important source of human salmonellosis at EU level, with 42.4% (7,903,000 cases, 95% Credibility Interval (CI) 4,181,000 – 14,510,000) of cases, followed by 31.1% attributed to pigs (5,800,000 cases, 95% CI 2,973,000 – 11,100,000). Broilers and turkeys were estimated to be less important sources of *Salmonella*, contributing with 12.6% (2,350,000 cases, 95% CI 736,300 – 6,194,000) and 3.8% (702,400 cases, 95% CI 325,500 – 1,590,000), respectively. A total of 1.6% (292,400 cases, 95% CI 150,700 – 562,700) of all salmonellosis cases were reported as being travel-related, and 0.1% (13,848) of cases were reported as being part of outbreaks with unknown source. *S.* Enteritidis was the most important serovar in the study, and of all infections caused by this serovar, 63% (7,504,000 cases, 95% CI 3,964,000-13,770,000) were attributed to laying hens, whereas 90.8% of *S.* Typhimurium originated from pigs (2,950,000 cases, 95% CI 1,510,000-

5,663,000). Country-specific results show laying hens as the most important source of salmonellosis in 13 countries (Austria, Czech Republic, Estonia, Germany, Greece, Hungary, Latvia, Lithuania, Luxembourg, Slovenia, Slovakia, Spain and the United Kingdom), whereas pigs were the larger animal contributor in eight (Belgium, Cyprus, Finland, France, Ireland, Italy, Poland and Sweden). In Finland and Sweden the majority of *Salmonella* infections were estimated to be travel-related. Travel was also an important source in Ireland, the UK and Denmark, although to a lower extent. In the Netherlands, the proportion of disease attributed to layers and pigs were similar. In Denmark, the most important food-animal source was estimated to be turkeys, and broilers were the major source in Portugal. Countries estimated to be the main origin of the food sources causing salmonellosis cases in the EU were Poland, with 21.3% of cases (3,563,710 cases, 95% CI 911,750 – 10,818,900), followed by 18.4% from Spain (3,081,090 cases, 95% CI 898,170 – 9,056,800) and 14.5% from Portugal (2,422,142 cases, 95% CI 361,368 – 8,508,397) (Manuscript II).

Danish strategies for risk management of *Salmonella* in the farm-to-fork continuum include the routine application of a source attribution model to estimate the contribution of the major animal-food sources to human infections by *Salmonella* in Denmark. This model concept formed the basis for the model described in Manuscript II. As part of the validation process of the EU model, results for Denmark in the EU model were compared with the ones obtained using the Danish model in the same period (Manuscript III).The Danish model points to pork as the main animal source of human salmonellosis in the period (9.3% of cases), followed closely by table eggs (7.5% of cases) and broilers (4.7% of cases), while the EU model attributed 18.0% to pigs, 19.6% to turkeys, 10.1% to eggs and 3.5% to broilers. Travel-related cases constitute 30.6% in the Danish model and only 23.7% in the EU model. Cases that could not be attributed to any source corresponded to 16.7% in the Danish model and 18.3% in the European model. Discrepancies in numbers are explained by differences in model structure and basic assumptions: a) cases with no travel information in the Danish model are redistributed according to proportions observed in cases with full information; in the EU model, as some countries did not provide any information regarding travel prior to sickness, it had to be assumed that no information means no travel; b) the Danish model uses data subtyped to phage-type level, which allows for a more specific allocation of cases to the right sources as compared to the EU model; c) the larger number of sources in the Danish model allows for more options for specific allocation of cases, presumably resulting in a more correct distribution of cases among sources; d) the Danish model uses official data on amount of domestic and imported food items available for consumption in the country, but does not as opposed to the EU model take into account the amount imported from each country specifically, which probably results in an underestimation of the contribution from high prevalence countries as compared to the EU model; e) the EU attributed sporadic cases were multiplied by an underreporting factor, altering the balance between sporadic and outbreak-related cases (Manuscript III). All things considered, the two models rank three out of the four sources in a similar order and, while the EU model is considered useful for countries which cannot readily attain the level of detailing found in Denmark for monitoring and surveillance data, Denmark would benefit more from applying country-specific data than to adopt the results of the EU model.

The last chapter presents an alternative approach to obtain results for the Czech Republic, Norway, Bulgaria and Romania, the last two of which were excluded from the EU-model due to insufficient data. Using clustering techniques, 28 countries were grouped according to variables used to characterize them as to social and economic status, animal production characteristics and food consumption patterns. Where available, variables reflecting the occurrence of *Salmonella enterica* in humans and animals were also used. The results of the analyses were delivered to a panel of experts composed by foodborne disease epidemiologists and risk modelers, which were asked to provide attribution estimates for the aforementioned countries, based on their similarity to countries for which results were previously obtained. Experts were also asked to evaluate the method concerning its utility and applicability of results. Individual estimates were evaluated based on comparison with the Czech results, for which results based on the microbial subtyping model were available, but also in relation to uniformity of guesses and uncertainty intervals among different estimates from the same expert and among all experts in the panel. This evaluation resulted in five out of the seven respondents being maintained in the panel. Although the Czech Republic values obtained did not match the ones observed in the EU study, the order of importance of the animal sources was in agreement

between the two studies and there was also a consensus in the panel concerning that order. It is, therefore, believed that with some adjustments, this method may be useful for prioritizing targeted actions for *Salmonella* control in countries without sufficient data for a traditional approach. Further on, this method may be used to identify "surrogate countries" from where animal prevalence data can be "borrowed" and applied in the microbial subtyping approach in the aforementioned Member States.

This PhD project has provided results for a European "source of infection account" for *Salmonella*, and has at the same time been evaluating the approaches attempted, raising questions and proposing solutions on how to deal with the lack of good-quality data for such studies. The project has also achieved results that may lay the groundwork for future attempts to develop *Salmonella* source attribution estimates in a more global perspective.

Sammendrag (Summary in Danish)

Denne PhD afhandling beskriver udviklingen af en matematisk model, der estimerer det kvantitative bidrag fra fire husdyrreservoirer til forekomsten af *Salmonella* infektioner hos mennesker i den Europæiske Union (Del I). Med henblik på at ekstrapolere resultaterne til lande, hvor datatilgængeligheden er mindre, præsenteres desuden en alternativ og mere udforskende metode baseret på ekspertvurderinger. Sidstnævnte skal ses som et første skridt på vejen til at kvantificere betydningen af forskellige smittekilder for human salmonellose i et mere globalt perspektiv (Del II).

Tilfælde af *Salmonella* infektioner blev tilskrevet udlandsrejse, fødevarebårne udbrud samt de fire husdyrreservoirer: svin, slagtekyllinger, kalkuner og konsumægsproducerende høner. Modellen inkluderede data fra 24 lande. Metoden kræver data vedr. *Salmonella* forekomst og serotypefordeling i husdyr, rapporterede infektioner hos mennesker, oplysninger om mulig erhvervelse af infektionen i udlandet (herfra benævnt "rejseinformation"), forekomst af fødevarebårne udbrud samt kilderne til disse, samt mængden af kød eller æg, som er til rådighed for forbrugerne i de enkelte lande. Datahåndtering, -analyse og -validering vurderedes at være af stor betydning for resultaternes kvalitet, og der blev derfor lagt vægt på at beskrive, hvad der kræves for at frembringe et datasæt med standardiserede oplysninger for alle lande (Manuskript I).

Data om rapporterede *Salmonella* infektioner hos mennesker blev skaffet fra det Europæiske Center for Sygdomsforebyggelse og Kontrol (ECDC) via Den Europæiske Fødevaresikkerhedsautoritet (EFSA). *Salmonella* forekomsten i de fire husdyrarter blev indhentet fra de EU-dækkende baseline undersøgelser (BS) (svin og kalkuner) rapporteret af EFSA, samt fra resultaterne af den harmoniserede overvågning af Salmonella (slagtekyllinger og æglæggere) rapporteret i Zoonoserapport (EUSR) ligeledes publiceret af EFSA. Oplysninger om fødevarebårne *Salmonella* udbrud blev også leveret af EFSA. Mængden af animalske fødevarer, der var til rådighed til forbrug blev estimeret på baggrund af data om fødevareproduktion og samhandel indhentet fra det europæiske statistisk kontor, EUROSTAT. Disse data blev suppleret med oplysninger fra sammenslutningen af fjerkræslagterier i EU, AVEC. Der var visse begrænsninger i data, som for nogle lande inkluderede manglende deltagelse i en eller flere af baseline undersøgelserne, manglende indberetning af fødevarebårne udbrud eller rejseinformation, manglende indberetning af serotype specifikke data, manglende indberetning af case-baserede data og manglende tilgængelighed af data i EUROSTAT. For at standardisere de foreliggende oplysninger, blev det antaget, at human tilfælde uden rejseinformation var indenlandsk erhvervede. Human tilfælde uden specifik serotype information blev tildelt en serotype i forhold til serotypefordelinger observeret i det samme datasæt eller fra andre referencedata. Manglende EUROSTAT data blev estimeret på baggrund af tidligere års data, og manglende data fra BS blev, hvor det var muligt, erstattet af data fra EUSR. For nogle lande var datamængden og – kvaliteten for ringe til, at de kunne indgå i modellen uden at kompromittere validiteten af resultaterne. Det endelige datasæt omfattede Østrig, Belgien, Cypern, Tjekkiet, Danmark, Estland, Finland, Frankrig, Tyskland, Grækenland, Ungarn, Irland, Italien, Letland, Litauen, Luxembourg, Holland, Polen, Portugal, Slovakiet, Slovenien, Spanien, Sverige og Storbritannien. Tre lande blev inkluderet i den indledende analyse, men ikke i det endelige datasæt. Det var Bulgarien, hvor 100% af de humane tilfælde ikke havde oplysning om serotype; samt Rumænien, som kun deltog i én baseline undersøgelse og ikke havde andre relevante data, og desuden havde en stor andel human tilfælde uden serotypeoplysning (Manuskript I).

Den Bayesiansk model, som blev anvendt til den matematiske analyse, sammenligner serotypefordelingen i mennesker med serotypefordelingen i husdyrreservoirerne. Modellen estimerer antallet af tilfælde af human salmonellose i de 24 lande fra hvert af disse reservoirer, samt fra rejser og udbrud baseret på de ovennævnte data (Manuskript II). Resultaterne viste, at æglæggere var den mest betydningsfulde kilde til human salmonellose i EU, ansvarlig for 42,4% (7.903.000 tilfælde, 95% Credibility Interval (CI) 4.181.000 – 14.510.000) af tilfældene, efterfulgt af svin med 31,1% af tilfældene (5.800.000 cases, 95% CI 2.973.000 – 11.100.000). Slagtekyllinger og kalkuner blev vurderet til at være mindre betydningsfulde kilder og bidrog med hhv. 12,6% (2.350.000 tilfælde, 95% CI 736.300 – 6.194.000) og 3,8% (702.400 cases, 95% CI 325.500 – 1.590.000) af tilfældene. I alt 1,6% (292.400 tilfælde, 95% CI 150.700 – 562.700) af alle salmonellose tilfælde blev rapporteret som værende rejserelaterede, mens 0,1%

(13.848) af tilfældene blev rapporteret som dele af udbrud med ukendt kilde. *S*. Enteritidis var den hyppigst forekommende serotype, hvoraf 63% (7.504.000 cases, 95% CI 3.964.000-13.770.000) af disse infektioner blev tilskrevet æglæggere, mens 90,8% af *S*. Typhimurium infektionerne blev estimeret til at komme fra svin (2.950.000 tilfælde, 95% CI 1.510.000-5.663.000). Landespecifikke resultater viste, at æg var den vigtigste kilde til salmonellose i 13 lande (Østrig, Tjekkiet, Estland, Tyskland, Grækenland, Ungarn, Letland, Litauen, Luxembourg, Slovenien, Slovakiet, Spanien og Storbritannien), mens svin var den store bidragyder i otte (Belgien, Cypern, Finland, Frankrig, Irland, Italien, Polen og Sverige). I Finland og Sverige kunne hovedparten af salmonellainfektionerne relateres til udlandsrejse. Rejse var også en vigtig kilde i Irland, Storbritannien og Danmark om end i lavere grad. I Holland var andelen af infektioner fra æg og svin omtrent det samme. I Danmark blev den vigtigste fødevarekilde estimeret til at være kalkun, mens slagtekyllinger var den største kilde i Portugal. Kigger man på det samlede bidrag fra de enkelte lande, blev kilder fra Polen vurderet til at bidrage med den største del af samtlige salmonellainfektioner med et estimat på 21.3% af tilfælde (3,563,710 cases, 95% CI 911,750 – 10,818,900). Polllen blev efterfulgt at Spanien med 18.4% (3,081,090 tilfælde, 95% CI 898,170 – 9,056,800) og Portugal med14.5% (2,422,142 tilfælde, 95% CI 361,368 – 8,508,397) . (Manuskript II).

Danske strategier for risikohåndtering af *Salmonella* i jord-til-bord kæden omfatter anvendelse af en såkaldt smittekilderegnskabsmodel, det estimerer bidraget fra de vigtigste animalske fødevarekilder til infektioner hos mennesker i Danmark. Det danske modelkoncept dannede grundlag for EU-modellen beskrevet i Manuskript II. Som en del af valideringsprocessen af EU-modellen, blev resultaterne for Danmark i EU-modellen sammenlignet med dem, der blev estimeret under brug af den danske model i samme periode (Manuskript III). Den danske model pegede på svinekød (9,3% af tilfælde), som den vigtigste kilde til salmonellose i perioden, efterfulgt af æg (7,5% af tilfælde) og slagtekyllinger (4,7% af tilfælde), mens EU modellen tilskrev 18,0% af tilfældene til svin, 19,6% til kalkuner 10,1% til æg og 3,5% til slagtekyllinger. Rejserelaterede tilfælde udgjorde 30,6% i den danske model og kun 23.7% i EU modellen. Tilfælde der ikke kunne relateres til nogen kendt kilde udgjorde 16,7% i den danske model og 18,3% i den europæiske model. De observerede uoverensstemmelser kan forklares ved forskelle i modellernes struktur og de grundlæggende antagelser: a) en andel af tilfældene uden rejseinformation tilskrives rejse i den danske model, hvilket baseres på proportionen af rejsetilfælde observeret for tilfælde med fuld rejseinformation; i EU-modellen antages det, at ingen rejseinformation er lig med ingen rejserelation, da mange lande ikke skelner mellem ”nej til rejse” og ”ingen rejseinformation”, b) den danske model anvender *Salmonella* typefordelinger baseret på både serotypning, fagtypning og resistensbestemmelse, hvilket giver en mere specifik fordeling af tilfælde til de rigtige kilder sammenlignet med EU modellen; c) et større antal smittekilder i den danske model giver flere muligheder for specifik fordeling af tilfælde, hvilket formentlig resulterer i en mere korrekt kildetildeling; d) den danske model anvender officielle data om mængden af dansk producerede og importerede fødevarer til rådighed til forbrug, men tager i modsætning til EU-modellen ikke hensyn til den mængde, der importeres specifikt fra hvert land, hvilket sandsynligvis resulterer i at bidrag fra lande med høje *Salmonella* forekomster underestimeres i den danske model; e) de sporadiske tilfælde i EU modellen blev multipliceret med en underrapporteringsfaktor, hvilket ændrer det relative forhold mellem sporadiske og udbrudsrelaterede tilfælde (Manuskript III). Alt taget i betragtning, så rangerer de to modeller tre ud af de fire kilder i samme rækkefølge, og mens EU-modellen må anses for at være nyttig for lande, som ikke umiddelbart har den datadetaljeringsgrad som findes i Danmark, vil Danmark kunne drage større nytte af at anvende landespecifikke importdata frem for at anvende resultaterne fra EU-modellen.

Det sidste kapitel beskriver en alternativ metode til at estimere kilder til human salmonellose for Tjekkiet, Norge, Bulgarien og Rumænien, hvoraf de to sidstnævnte ikke var inkluderet i EU modellen pga. manglende data. Ved hjælp af clusteranalyser blev 28 lande grupperet efter nogle udvalgte variable, som karakteriserede landenes sociale og økonomiske status, den animalske husdyrproduktion samt kostvaner. Hvis data var tilgængelige, blev variable som afspejler forekomsten af *Salmonella* hos mennesker og husdyr også inddraget. Resultaterne af analyserne blev fremlagt et ekspertpanel med speciale indenfor fødevaresikkerhed, epidemiologi og risikomodellering. Disse blev bedt om at komme med estimater for den relative betydning af *Salmonella* smittekilder for de førnævnte lande, baseret på disses lighed med lande, for

hvilke resultater allerede forelå dvs. på baggrund af resultater fra EU-modellen. Eksperterne blev også bedt om at evaluere metodens egnethed og anvendeligheden af resultaterne. Eksperternes individuelle estimater blev evalueret dels ved en sammenligning med de tjekkiske resultater, som var til rådighed fra EU-modellen, men også i forhold til estimaternes ensartethed og usikkerhedsintervallerne mellem de forskellige estimater fra samme ekspert og imellem eksperterne i panelet. Evalueringen resulterede i, at svarene fra fem ud af de syv respondenter blev bibeholdt i de endelige analyser. Selv om panelet angav estimater for Tjekkiet som ikke var identiske med dem fra EU-modellen, var der enighed om rækkefølgen af betydningen af de animalske kilder, og der var enighed i panelet om samme rækkefølge. Det vurderes derfor, at metoden med nogle justeringer, kan være nyttig til at prioritere målrettet *Salmonella* kontrol i lande uden tilstrækkelige data til at gennemføre en mere datadrevet fremgangsmåde. På sigt kan metoden måske bruges til at identificere "surrogatlande," hvorfra prævalensdata kan "lånes" og anvendes i en matematisk model baseret på sammenligning af *Salmonella* typer.

Dette PhD projekt har fremlagt resultater for et Europæisk smittekilderegnskab for *Salmonella*, samt evalueret de anvendte metoder og fremkommet med løsninger til, hvordan man kan håndtere manglende eller utilstrækkelige data i lignende undersøgelser. Projektet har også opnået resultater, som kan lægge grunden for fremtidige forsøg på at udarbejde *Salmonella* smittekilderegnskaber i et mere globalt perspektiv.

Resumo (Summary in Portuguese)

Esta tese apresenta um modelo matemático usado para estimar a contribuição de quatro reservatórios animais da cadeia de produção de alimentos para o número de casos humanos de salmonelose na União Européia (UE) (Parte I). Além disso, uma abordagem alternativa baseada na opinião de peritos foi testada como ferramenta para extrapolar resultados para países com menor disponibilidade de dados, o que respresenta um primeiro passo para atribuição de *Salmonella* em nível mundial (Parte II).

Casos de salmonelose em humanos foram atribuídos a viagens, surtos de doenças de transmissão alimentar e quatro reservatórios animais da cadeia de produção de alimentos (suínos, frangos de corte, perus e poedeiras) usando um modelo Bayesiano baseado em subtipagem microbiana em 24 países da UE. O método escolhido requer uma grande quantidade de dados, como prevalência de *Salmonella* em animais, casos notificados em humanos, possibilidade de infecção no exterior (daqui para frente referido como "histórico de viagem"), casos em humanos relacionados a surtos e quantidade de carne ou ovos de cada reservatório animal que é produzida em um país e se encontra disponível para consumo nos outros. Por esse motivo, a preparação de um banco de dados com informações padronizadas para todos os países requereu um manuseio específico dos dados (Manuscrito I).

Dados sobre casos esporádicos foram fornecidos pelo European Centre for Disease Prevention and Control (ECDC) através da European Food Safety Authority (EFSA), que também forneceu os dados de surtos. Prevalências de *Salmonella* em animais foram retiradas dos estudos de nível-base (BS) (suínos e perus) conduzidos pela EFSA entre 2004 e 2008 e completados onde necessário com dados da vigilgância padronizada da União Européia (frangos de corte e poedeiras) encontradas nos European Union Summary Report (EUSR), também publicados pela EFSA. O volume de alimento disponível para consumo foi calculado com base nos dados de comércio retirados do European Statistical Office (EUROSTAT), e completados com informações da Association of Poultry Processors and Poultry Trade in the European Union Countries (AVEC). Limitações encontradas incluem a não-participação de alguns países em todos os BS, a não-notificação de casos individualmente, não-notificação específica de sorovares e indisponibilidade de registros de comércio no EUROSTAT. Para padronizar as informações disponíveis, todos os casos sem histórico de viagem foram considerados como infecção no próprio país; casos sem identificação apropriada até o nível de sorovar foram re-classificados de acordo com as proporções de sorovares existentes no banco ou em outros documentos de referência; informações de comércio não encontradas foram estimadas com base nos anos para os quais os dados estavam disponíveis, e a não-participação nos BS foi subtituída, quando possível, com dados de vigilância dos EUSR. Países em que a falta de dados foi considerada extrema a ponto de ameaçar os resultados do modelo foram excluídos. O banco resultante contém dados da Austria, Belgica, Chipre, República Tcheca, Dinamarca, Estônia, Finlândia, França, Alemanha, Grécia, Hungria, Irlanda, Itália, Letônia, Lituânia, Luxemburgo, Holanda, Polônia, Portugal, Eslováquia, Eslovênia, Espanha, Suécia e Reino Unido. Três países foram incluídos nas análises preliminares, mas foram retirados da lista final: a Bulgaria, que notificou 100% dos casos sem detalhamento de sorovares; e a Romênia, que só participou de um BS e não tinha dados suficientes publicados no EUSR (Manuscrito II).

Um modelo Bayesiano que compara a presença de sorovares em humanos com a presença dos mesmos sorovares em animais da cadeia de produção de alimentos foi aplicado para estimar a contribuição de cada uma dessas categorias animais para o número de casos de salmonelose nos 24 países incluídos no banco descrito anteriormente (Manuscript II). Galinhas poedeiras (i.e., ovos) foram consideradas a fonte mais importante de salmonelose na União Européia, com 48.1% 42.4% (7,903,000 casos, Intervalo de Credibilidade de 95% (IC) 4,181,000 – 14,510,000) dos casos, seguidas de 31.1% attribuídos a suínos (5,800,000 casos, IC 95% 2,973,000 – 11,100,000). Perus e frangos de corte foram considerados fontes de menor importância, contribuindo com 12.6% (2,350,000 casos, IC 95% 736,300 – 6,194,000) e 3.8% (702,400 casos, IC 95% 325,500 – 1,590,000), respectivamente. Um total de 10.2% de todos os casos de salmonelose esteve relacionado a viagens, e 3.9% dos casos foram parte de surtos sem identificação do alimento implicado. *S.* Enteritidis foi o sorovar mais importante no estudo, tendo sido responsável por 95.9% dos casos atribuídos a poedeiras, 56.9% dos casos atribuídos a frangos de corte, 30.4% dos atribuídos a perus

e 28.3% dos casos atribuídos a suínos, nos quais o sorovar mais importante foi *S.* Typhimurium (63.1% dos casos atribuídos a essa fonte). Os resultados por país demonstraram que poedeiras são a fonte mais importante de salmonelose em 13 países (Áustria, República Tcheca, Estônia, Alemanha, Grécia, Hungria, Letônia, Lituânia, Luxemburgo, Eslovênia, Eslováquia, Espanha e Reino Unido), enquanto suínos foram o maior contribuinte em oito (Bélgica, Chipre, Finlândia, Franca, Irlanda, Itália, Polônia e Suécia), apesar de na Finlândia e Suécia a maior parte dos casos ter origem no exterior. Na Holanda, a proporção de casos atribuídos a poedeiras e suínos foi parecida. Na Dinamarca, o reservatório animal mais importante foram os perus, e frangos de corte foram a fonte mais importante em Portugal. Viagens ao exterior também foram uma fonte importante, apesar de menos que nos países citados anteriormente, na Irlanda, Reino Unido e Dinamarca. Os países que foram estimados como principais origens dos reservatórios de salmonelose na UE foram a Polônia, com 21.3% dos casos (3,563,710 casos, IC 95% 911,750 – 10,818,900), seguida de 18.4% da Espanha (3,081,090 casos, IC 95% 898,170 – 9,056,800) e 14.5% de Portugal (2,422,142 casos, IC 95% 361,368 – 8,508,397) (Manuscrito II).

As estratégias de controle de risco em salmonelose na Dinamarca incluem a aplicação periódica de um modelo de atribuição para estimar a contribuição dos principais animais e alimentos para casos humanos de salmonelose no país. Como parte do processo de validação do modelo europeu, seus resultados foram comparados com os do modelo dinamarquês (Manuscript III). O modelo dinamarquês tem os suínos como a principal fonte de salmonelose no periodo estudado (9.3%), seguido de ovos (7.5%) e frangos de corte (4.7%), enquanto o modelo europeu atribuiu 18.0% a suínos, 19.6% a perus, 10.1% a poedeiras e 3.5% a frangos de corte. Casos relacionados a viagens ao exterior corresponderam a 30.6% no modelo dinamarquês e apenas 23.7%. Casos que não foram atribuídos a nenhuma fonte foram 16.7% no modelo dinanarquês e 18.3% no europeu. As diferenças nos números observados são explicadas por diferenças na estrutura dos modelos e em seus pressupostos básicos: a) casos sem histórico de viagem no modelo dinamarquês são redistribuídos de acordo com as proporções observadas nos casos com informação completa; no modelo europeu, como alguns países não possuíam informação nenhuma a respeito de viagens, foi necessário pressupôr que casos sem histórico de viagem não viajaram; b) o modelo dinamarquês usa dados subtipados até o nível de fagotipos, que permite a alocação mais específica de casos às fontes corretas, se comparado ao modelo; c) a maior variedade de alimentos e animais no modelo dinamarquês oferece mais opções para a atribuição específica de casos, de forma que menos casos são direcionados à categoria "fonte desconhecida"; d) o modelo dinamarquês usa dados oficiais nacionais sobre o volume de alimentos nacionais e importados disponíveis para consumo no país, não levando em consideração o volume importado de cada país de origem; isto resulta na contribuição específica de países com altas prevalências sendo "diluídas" no total importado, enquanto no modelo europeu a combinação da prevalência com o volume de importações tem mais impacto nos resultados e) os casos esporádicos atribuídos no modelo da UE foram multiplicados por um fator de correcão de subnotificacão, alterando o equilíbrio entre casos esporádicos e casos ligados a surtos (Manuscrito III). Tendo em mente todas as observações feitas, os dois modelos ordenam três das quarto fontes em ordem de prioridade semelhante e, apesar de o modelo europeu ser considerado útil para países que não possuem o mesmo nível de detalhamento de dados de vigilância que a Dinamarca, este país ganharia mais adaptando o modelo atual para usar dados de comércio entre países que adotando o modelo europeu.

O último capítulo apresenta uma abordagem alternativa para obter resultados de atribuição na Republica Tcheca, Bulgária, Noruega e Romênia. Usando técnicas de cluster analysis, 28 países foram agrupados de acordo com variáveis usadas para caracterizá-los do ponto de vista sócio-econômico, de produção animal, clima e hábitos de consumo de alimentos. Quando disponíveis, dados de *Salmonella* em humanos e animais também foram incluídos. O resultado das análises foi distribuído a um painel de especialistas em segurança alimentar, epidemiologia e modelagem de risco, e foi pedido ao painel que estimasse resultados de atribuição para os países citados anteriormente com base em suas semelhancas com outros países. Também foi pedido que o método fosse avaliado em termos de utilidade e aplicabilidade dos resultados, e as estimativas dos especialistas foram avaliadas através da comparação com os valores obtidos para a República Tcheca no modelo europeu. Essa avaliação resultou na permanência de cinco dos sete especialistas originais no painel. Apesar de os valores específicos obtidos por esse método serem diferentes

dos do modelo europeu, a ordem de prioridade entre os reservatórios animais foi a mesma. Portanto, é possível que, com adaptacões, esse método possa ser útil para ajudar a definir prioridades de ação no controle de *Salmonella* em países que não possuem dados suficientes para uma abordagem mais tradicional. O método também pode ser usado para definir "países substitutos", dos quais os resultados dos métodos mais tradicionais possam ser copiados, dadas as semelhanças entre países.

O estudo de doutorado aqui apresentado obteve resultados que estabelecem parte das fundações para estudos de atribuição de fonte em nível global, avaliando, ao mesmo tempo, os métodos testados e propondo soluções para lidar com dificuldades relacionadas à má qualidade potencial dos dados disponíveis para esses estudos.

List of abbreviations

AVEC:	Association of Poultry Processors and Poultry Trade in the EU Countries
BS:	EFSA Studies on the baseline prevalence of *Salmonella* in animal sources in the European Union
CDB:	Country Data Bank
DTU:	Technical University of Denmark
EC:	The European Commission
ECDC:	European Center for Disease Prevention and Control
EFSA:	European Food Safety Authority
EU:	European Union
EUROSTAT:	The statistical office of the European Union
EUSR:	The European Union Summary Report on Trends and Sources of Zoonoses, Zoonotic Agents and Food-borne Outbreaks
FAO:	Food and Agriculture Organization
FERG:	Foodborne Disease Burden Epidemiology Reference Group
GEMS/Food:	Global Environment Monitoring System Food Consumption Cluster Diets
GFN:	Global Foodborne Infections Network
MCMC:	Markov Chain Monte Carlo
MS:	Member State
OIE:	World Organization for Animal Health
PSI:	Proportional Similarity Index
SSI:	Statens Serum Institute
TESSy:	The European Surveillance System
UF:	Underreporting Factor
UNDP:	United Nations Development Program
WHO:	World Health Organization

Country abbreviations

AT: Austria
BE: Belgium
BG: Bulgaria
CH: Switzerland
CY: Cyprus
CZ: Czech Republic
DE: Germany
DK: Denmark
EE: Estonia
ES: Spain
FI: Finland
FR: France
GR: Greece
HU: Hungary
IE: Ireland
IT: Italy
LT: Lithuania
LU: Luxembourg
LV: Latvia
MT: Malta
NL: The Netherlands
NO: Norway
PL: Poland
PT: Portugal
RO: Romania
SE: Sweden
SI: Slovenia
SK: Slovakia
UK: United Kingdom
IE: Ireland

1. TABLE OF CONTENTS

2. OUTLINE

This thesis is divided into two parts:

Part I presents a microbial subtyping approach to attribute cases of human salmonellosis to animal reservoirs in 24 Member States of the European Union. The necessary data management to create a dataset with standardized information from the countries finally included in the model is also described. A comparison of Danish *Salmonella* source attribution estimates achieved by the developed EU model and a Danish model used routinely as part of the national risk management activities is presented to discuss the validity of both models. The results from Part I form the background for the methodology in Part II.

Part II presents the pilot of a novel approach to use clustering techniques and expert elicitation to extrapolate estimates from the study described in Part I to European countries with insufficient *Salmonella* data. The thesis closes with a general discussion and conclusions.

3. INTRODUCTION

Foodborne infections are widespread and a growing public health problem worldwide. Recent increases in international trade, migration and travel provide an opportunity for pathogens to spread at a much faster rate and to a much larger area than in previous decades (Greger, 2007; Tauxe et al., 2010). As an example, it is estimated that, in the United States, foodborne diseases caused by the 31 major known pathogens result in 9.4 million cases, 55,961 hospitalizations and 1,351 deaths each year (Scallan et al., 2011a). In addition, unspecified foodborne agents (pathogens with insufficient information for agent-specific calculations, known agents not yet recognized as causing foodborne illness and chemicals or other substances present in food but with pathogenicity not yet proven) are estimated to cause 38.4 million cases per year, resulting in 71,878 hospitalizations and 1,686 deaths (Scallan et al., 2007b). Globally, although the true burden of foodborne diseases is currently unknown, it is estimated that food- or waterborne diarrheal diseases are responsible for 2.2 million deaths per year worldwide, 1.9 million of which are children (WHO, 2007).

Salmonella enterica is considered one of the leading causes of gastroenteritis and bacteremia in the world (Scallan et al., 2011a, Hendriksen et al., 2011), being estimated to cause 93.8 million human cases and 155 thousand deaths every year (Majowicz et al., 2010).

The genus *Salmonella* consists of only two species, namely *Salmonella enterica* and *Salmonella bongori* (WHOCC-Salm, 2007). *Salmonella enterica* is divided in six subspecies (*S. enterica enterica, S. enterica salamae, S. enterica arizonae, S. enterica diarizonae, S. enterica houtenae* and *S. enterica indica*) (Tindall et al., 2005, Haesebrouck et al., 2005), of which *S. enterica enterica* and *S. enterica salamae* are commonly found in warm-blooded animals, while the others are more frequent in cold-blooded animals and in the environment (WHOCC-Salm 2007).

Although other sources are recognized (Baker et al., 2007; O'Reilly et al., 2007), transmission of *Salmonella* to humans occurs mainly through the ingestion of contaminated food (Acha and Szyfres, 2001; EFSA, 2011a). Implicated foods are frequently beef, pork, poultry, dairy products, eggs and fresh produce, and scientific evidence confirms the transmission of strains from the animal reservoir through the food chain and to the human population (Newell et al., 2010).

Identifying which foods are more frequently implicated in the transmission of an illness is a crucial step on the prioritization of control activities (Kuchenmüller et al., 2009). This process is called *source attribution*, and it can be based on different approaches, such as analysis of outbreak data, analysis of sporadic cases, microbial subtyping, comparative exposure assessment, intervention studies and expert elicitations (Pires et al., 2009). Methods for source attribution are intended to provide tools for the setting of priorities in relation to human foodborne and zoonotic diseases, being a critical tool for decision-making aimed at reducing human infections faster and more effectively (Havelaar et al., 2007).

In 2006, the World Health Organization (WHO) created the Foodborne Disease Epidemiology Reference Group (FERG) as part of a strategy to estimate the global burden of foodborne diseases. The group is organized in thematic task forces, one of which is focused on attributing illnesses to food sources (WHO, 2009). This is an ongoing process, which has not yet been accomplished, mainly because nationally representative prevalence data about foodborne pathogens in humans, animals and food items are not available in most parts of the world.

Countries which have a more favorable data situation have attempted to estimate the contribution of the major animal-food sources to human infections of foodborne pathogens. A part of these efforts have been directed towards *Salmonella*, like in Denmark (Hald et al., 2004; Pires and Hald, 2010), Sweden (Whalström et al., 2011), Japan (Toyofuku et al., 2011), New Zealand (Müllner et al., 2009) and the United States (Guo et al., 2011). A EU-wide source attribution approach based on outbreak data was also developed (Pires et al., 2010); this model attributed disease at the EU region level and did not provide estimates at country level. Also due to model limitations, achieved results were found insufficient to inform risk management decisions (see limitations of the study in Pires et al., 2010).

Since 2003, efforts have been made in the EU to standardize the reporting of pathogens and diseases in humans and animals[1]. Part of those efforts was the conduction of EU-wide studies to estimate the baseline prevalence of *Salmonella* in laying hens, slaughter pigs, turkeys and broilers in EU Member States, and the use of those data to set targets for reduction of *Salmonella* in those populations. More recent efforst include the harmonization of the monitoring of *Salmonella* in laying hens (in 2006), broilers and turkeys (not implemented until after the completion of this thesis). Those actions are expected to have an impact on the relative contribution of different food-animals to human salmonellosis in all individual MS, but until 2009, this information had not been assessed. This prompted the European Food Safety Authority (EFSA) to issue a procurement procedure, requiring the estimation of the relative contribution of different food and animal sources to *Salmonella* infections in humans in the EU and European regions (Question No EFSA-Q-2010-00685), using data sources officially approved and validated by the EU. This thesis presents the methods and results of the PhD project developed in response to that requirement, under contract number CT/EFSA/Zoonoses/2010/02 signed between EFSA and the National Food Institute, Technical University of Denmark.

3.1. *Salmonella* in humans

The typical clinical signs and symptoms of non-typhoidal salmonellosis in humans are acute fever, abdominal pain, nausea and vomiting, after an incubation period of 6-72 hours. Most infections are self-limiting, lasting about two to four days, and symptoms are often mild, with dehydration as the main serious feature. Extra-intestinal infection is not common, but when it happens, particularly in bloodstream infections, the disease can be life-threatening. A small percentage of convalescents can act as healthy carriers for weeks or months, and sometimes chronic sequelae, such as reactive arthritis, may follow recovery (Acha and Szyfres, 2001).

In the European Union (EU) in 2009, a total of human 109,844 cases were reported by 27 EU Member States (MS), most of which by serovars Enteritidis (52.3%), Typhimurium (23.3%) and Infantis (1.6%). Other serovars were present, but each one was detected in less than one percent of cases, adding up to 22.8%. Of the total cases reported, 108,614 were confirmed by laboratory, corresponding to a notification rate of 23.7 cases per 100,000 population, and showing a decrease when compared to previous years (EFSA, 2011a) (Figure 1).

[1] OJ L 268, 3.10.1998, p. 1–7. Decision No 2119/98/EC of the European Parliament and of the Council of 24 September 1998 setting up a network for the epidemiological surveillance and control of communicable diseases in the Community.

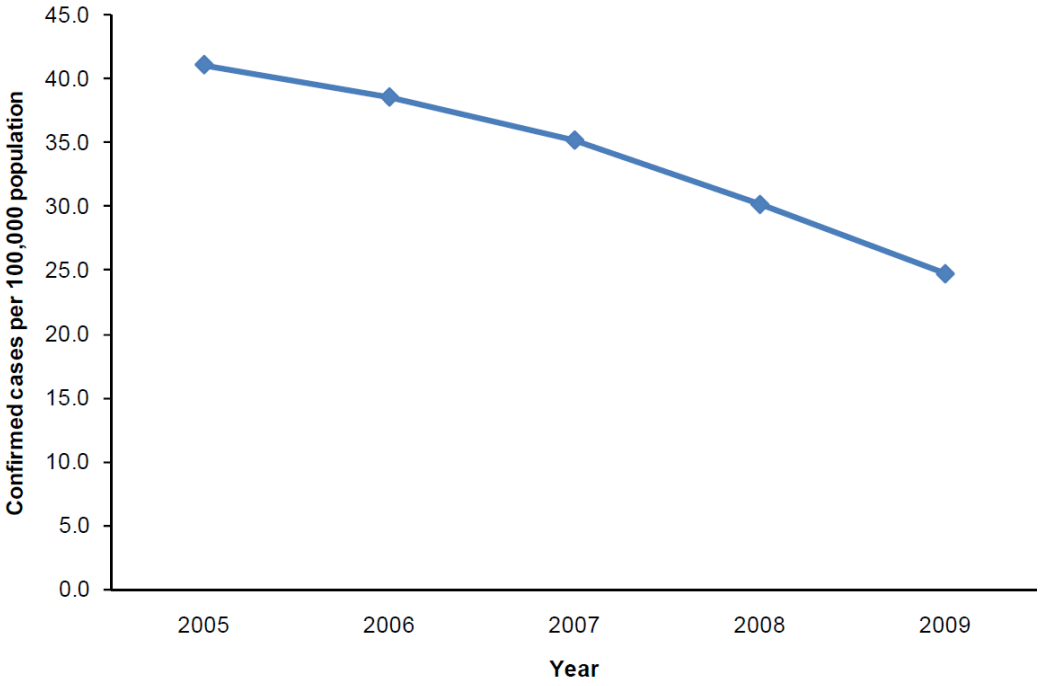


Figure 1. Notification rates of human salmonellosis in 2005-2009 by 25 EU Member States (AT, BE, CY, CZ, DE, DK, EE, ES, FI, FR, GR, HU, IE, IT, LT, LU, LV, MT, NL, PL, PT, SE, SK, SI, UK). Source: The European Union Summary Report on Trends and Sources of Zoonoses, Zoonotic Agents and Food-borne Outbreaks in 2009 (EFSA, 2011a).

One of the main obstacles for the evaluation of these numbers is the underreporting of cases, which can happen at all levels of the reporting pyramid, as described by Wheeler et al. (1999) (Figure 2). In the reporting pyramid, the real (and generally unknown) number of illnesses occurring in the population are represented at the base, and the number of cases reported in the surveillance system are represented at the top. The difference between the two totals is explained by 1) the percentage of cases which seek medical care; 2) the percentage of those which are asked to submit clinical specimens and actually provide them; 3) the percentage of specimens which are tested; 4) the sensitivity of the laboratory tests used; 5) the percentage of positive results which are reported and 6) the percentage of records in the reporting system which have complete and valid data. It is generally understood that the number of cases found at the tip of the pyramid is considerably smaller than the one found at its bottom. It is, therefore, accepted that the true burden of human salmonellosis (and other gastrointestinal infections) may be considerably larger than the reported incidence. Also, the level of underreporting varies strongly between countries, depending on differences in organization and effectiveness of local surveillance systems (de Jong and Ekdahl, 2006; ECDC, 2007).

The percentage of cases lost between the steps of the pyramid can be assessed in a country, for example through population surveys, hospital surveys, review of clinical records and a survey or evaluation of the laboratories. Several attempts have been made in recent years to "correct" the values officially reported, for example in England (Wheeler et al., 1999), the United States (Voetsch et al., 2004) and the Netherlands (Havelaar et al., 2012a). Two studies have estimated underreporting factors for the European Union (de Jong and Ekdhal 2006; Havelaar et al.,2012b), based on the incidence of disease in Swedish travelers returning home from within-EU travel. A global-level study was also conducted by Majowicz et al. (2010), using a combination of population-based studies, studies which calculated underreporting factors, disease notification, traveler return data and extrapolation.

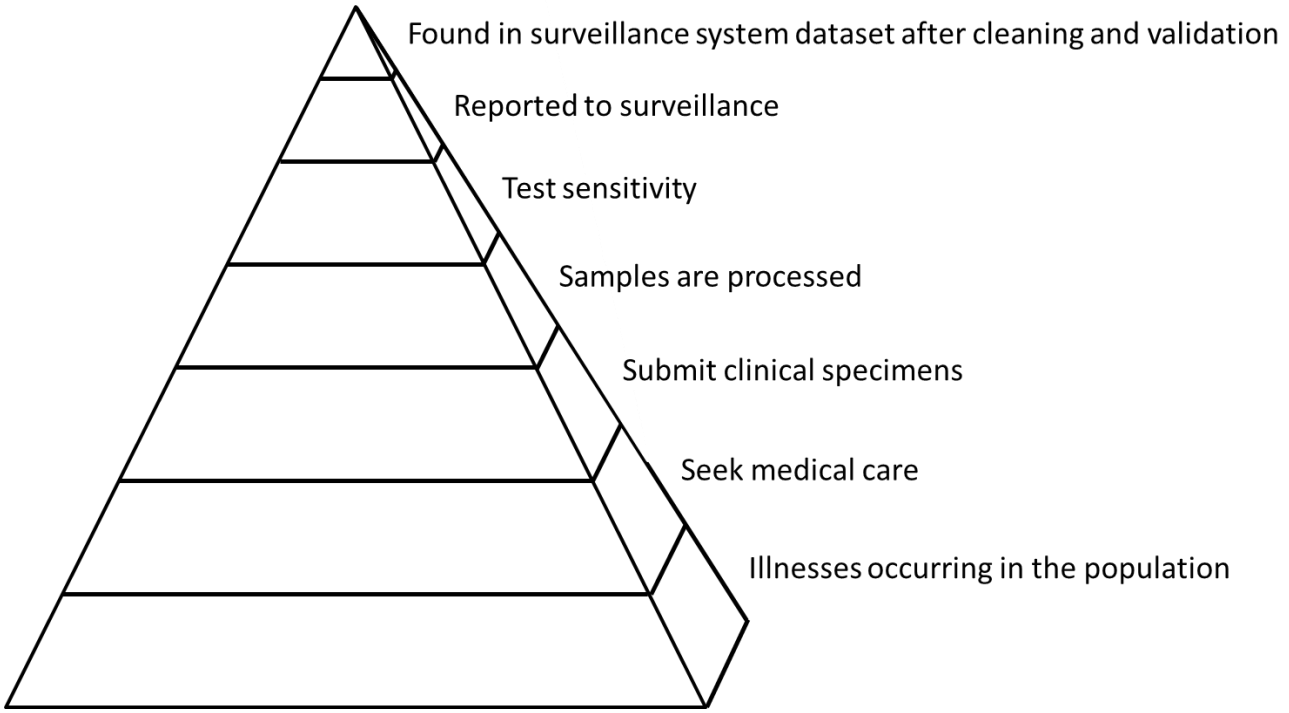


Figure 2. Reporting pyramid showing the steps where cases present in the population are "lost" by the surveillance system.

3.2. Sources of human salmonellosis

It is generally recognized that subtypes of *Salmonella* can be host-adapted, host-specific or host-ubiquitous, a concept that is based on epidemiological evidence (Kingsley and Bäumler, 2000). Contrary to what the name suggests, host-specific types are not found in only one species, but are only able to establish a stable population through direct contact in specific animal hosts, being still able to infect humans or other animals (Uzzau et al., 2000). When infecting other species, their pathogenicity is affected by their degree of host adaptation, generally causing a more severe clinical syndrome than in the original host population. *S.* Typhi and *S.* Paratyphi are adapted to man, where they cause severe systemic illness characterised by fever and abdominal symptoms (enteric/(para)typhoid fever) (Miller et al., 1995; Uzzau et al., 2000; Acha and Szyfres, 2001). These serovars are usually not pathogenic to animals and are not considered to have a zoonotic potential.

As mentioned earlier in this thesis, subtypes of *Salmonella enterica enterica* are mostly found in warm-blooded animals, meaning that over 1,500 subtypes of *Salmonella* are potentially pathogenic for humans. Non-typhoidal ubiquitous subtypes like *S.* Typhimurium affect humans and a wide range of animals, usually causing gastrointestinal infections of varying severity. There is also a group of serotypes that are highly adapted to an animal host, e.g. *S.* Cholerasuis in pigs, *S.* Dublin in cattle, *S.* Abortus-ovis in sheep and *S.* Gallinarum in poultry (Hald, 2011). These serovars only occasionally infect humans, where they may produce no, mild or serious disease (Acha and Szyfres, 2001; Mølbak et al., 2006).

In animals, most infection are sub-clinical, allowing transmission between herds or flocks and to humans before detection. Infected bovines may succumb to fever, diarrhea and abortion. Within calf herds,

Salmonella may cause outbreaks of diarrhea with high mortality. Fever and diarrhea are less common in pigs than in cattle, sheep and horses; goats usually show no signs of infection, and poultry generally develops serious illness only when infected with specific strains like *S.* Gallinarum and *S.* Pollurum (Acha and Szyfres, 2001). Although zoonotic *Salmonella* types can occur in almost all food-producing animals, even host-ubiquitous types are often more strongly associated to a particular animal reservoir. Thus, recognizing the main hosts of specific subtypes is of utmost importance for the identification of contaminated food sources (Hald, 2001).

It should be noted, though, that the potential of a food item as a transmission source depends not only on the infecting serovar, herd/flock prevalence or contamination at retail level, but also on the way it is traditionally prepared in the country of consumption. Food-preparation methods play an important part in human infection from contaminated food sources, as *Salmonella* optimally grows in temperatures around 37 degrees Celsius, and can be inactivated by thoroughly cooking the food (Adams and Moss, 1995; Kovats et al., 2004).

3.3. Source attribution methods and models

Source attribution of human illnesses can be defined as "the partitioning of the human disease burden of one or more foodborne infections to specific sources, where the term *source* includes animal reservoirs and vehicles (e.g. food)" (Pires et al., 2009). Methods present different advantages and limitations, and the applicability of each one depends on the public health question being addressed, on the characteristics of the pathogen and on data availability. Depending on the method chosen, attribution can be conducted at the point of reservoir, meaning at the origin of the pathogen, or at the point of exposure, i.e. at the end of the transmission chain.

A variety of data sources and analytic approaches can be used to attribute foodborne illnesses to food or animal sources (Batz et al., 2005; Pires et al., 2009), including:

- Microbiologic approaches: microbial subtyping and comparative exposure assessment.
- Epidemiologic approaches: analyses of case-control studies and analysis of outbreak data.
- Intervention studies.
- Expert elicitations.

Microbial subtyping approaches attribute disease at the reservoir level, while comparative exposure assessment and epidemiological studies focus on illness at the point of consumption, as outbreaks and case-control studies investigate the exposures common to all cases.

Most available methods have been applied to attribute human salmonellosis to its sources. Comparative exposure assessments are used to determine the relative importance of the known transmission routes of a hazard, by estimating the human exposure in each step of each possible route. It has been used for *Salmonella* in Denmark (Pires, 2009). *Salmonella* source attribution studies using data from outbreaks have been conducted in Europe (Pires et al., 2010; Pires et al., 2011a), Latin America and the Caribbean (Pires et al., 2012), Japan (Pires et al., 2011b), Canada (Ravel et al., 2009) and New Zealand (King et al., 2011). A source attribution study using a global meta-analysis of case-control studies of sporadic infections was recently published by Domingues et al. (2012). Expert elicitations have also been applied to *Salmonella* (Hoffmann et al., 2007a), and can also be used when the data required for a data-driven statistic approach are highly uncertain or unavailable, or when there is a need to fill data gaps or combine conflicting results from existing studies or approaches (Batz et al., 2005). Additionally, expert elicitations are particularly useful to

estimate the proportion of human cases that are attributable to each of the main transmission pathways: food- or waterborne, environmental, direct contact or person-to-person transmission. This type of attribution is not possible to attain by the direct application of the other methods described, as in neither of these, it is possible to include all possible routes of transmission. The delineation of the major transmission pathways therefore depends on the experts' critical analysis of the results of all relevant studies, in order to quantify the importance of each of the transmission pathways.

3.3.1. The microbial subtyping approach

Characterization of *Salmonella* subtypes beyond the subspecies level can be performed using phenotypic or genotypic approaches (Gebreyes et al., 2006). Phenotypic methods include serotyping, phage typing and characterization of antimicrobial resistance (AMR) profiles.

Serotyping by slide agglutination is based on the immunologic reactivity and antigenic determinants present in the cell surface, such as antigen O (membrane), K (capsule) and H (flagellar) (Haesebrouck et al., 2005). It is extensively used to categorize bacterial pathogens, and the serotyping scheme for the genus *Salmonella* is based on the Kauffman-White typing scheme (Gebreyes and Thakur, 2011). The serotyping of *Salmonella* generates a list of over 2,500 serovars, of which around 1,500 (60%) are subtypes of *S. enterica enterica* (WHOCC-Salm 2007). In some cases, subtyping is performed only based on the O-antigen, and is called serogrouping. Serotyping according to the Kauffmann-White scheme is the primary characterization method of *Salmonella*, being applied all over the world and harmonized to a degree that allows results to be compared between laboratories and countries (Baggesen et al., 2010).

It is also possible to characterize *Salmonella* isolates based on their susceptibility to specific bacteriophage viruses (Gebreyes et al., 2006). The method is independent from serotyping, and in theory, any *Salmonella* could be phage typed, as long as a specific panel of bacteriophages is obtained. However, due to their importance in humans, traditionally *Salmonella enterica enterica* ser. Enteritidis and *Salmonella enterica enterica* ser. Typhimurium (normally abbreviated as "*S.* Enteritidis" and "*S.* Typhimurium") are the ones phage typed, as a way to further and more specifically subtype isolates already serotyped. Given the specificity of phages to the target bacteria, this method provides a more discriminative characterization than serotyping, making it a better tool for detecting more specific relations between subtypes and hosts or food-sources (Mølbak and Neimann, 2002), which is of utmost importance, for example, during outbreak investigations (Hendriksen, 2010; Baggesen et al., 2010).

Antimicrobial resistance (AMR) profiling may be used to further subtype microorganisms based on their resistance to a panel of antimicrobials of different classes. In the last decades, it became a more common approach, as it also allows the identification of emergent drug-resistant strains (Gebreyes and Thakur, 2011).

Variations of the microbial subtyping approach using serotyping, phage typing and/or antimicrobial susceptibility profiling have been used for *Salmonella* in Denmark (Hald et al. 2004; Pires and Hald, 2010) and the Netherlands (van Pelt et al., 1999; Valkenburgh et al., 2007). The model used in Denmark has also been adapted to attribute salmonellosis in Japan (Toyofuku et al., 2011), Sweden (Whalström et al., 2011), the United States (Guo et al., 2011) and New Zealand (Mullner et al., 2009). Both the Dutch and the Danish model compare the number of reported human cases caused by a subtype with the relative occurrence of that subtype in the animal-food sources; the Dutch model assumes that the impact of each source is equal within each subtype, and that a source with a high relative occurrence of a type will necessarily result in more cases, ignoring that certain food types are traditionally prepared in ways that allow more or less survival of a

pathogenic load. For instance, a low prevalence of *S.* Enteritidis in table eggs usually cause more cases of human salmonellosis than a higher prevalence in broilers, because table eggs are more traditionally consumed raw or only lightly cooked. This feature is considered in the Danish model, which is described in section 3.3.1.1. The models by Toyofuku et al. (2011), Whalström et al. (2011) and Guo et al. (2011) – a.k.a. the CDC model – are all direct applications or adaptations of the Danish model, as described in their methods.

When it comes to the use of genotypic subtyping for source attribution, the Asymetric Island model has been used for *Campylobacter* in England (Wilson et al., 2008), New Zealand (Müllner et al., 2009) and Denmark (Boysen, 2012). Its use for *Salmonella*, particularly at EU-level, is still to be explored, since it uses Multil-Locus Sequence Typing (MLST) data, which are not available in most EU countries. Initiatives to use Multi-Loci VNTR Analysis (MLVA) data (which is based on the method as DNA-fingerprinting) for attribution purposes are ongoing (Pires and Hald, *pers. comm.*), but until the conclusion of this thesis, that had not yet been achieved.

In the EU, the reporting of *Salmonella* phage types, AMR profiles or DNA-based information are not required by Decision No 2119/98/EC[2] of the European Commission, and so the methods are only applied during outbreaks in most countries.

3.3.1.1. The Danish source account model, a.k.a. the Hald model

The DTU National Food Institute routinely applies a source attribution model to estimate the contribution of the major animal-food sources to human infections of *Salmonella*. The model was first implemented in 1995, and has since then evolved from being purely deterministic to becoming a stochastic model, built under a Bayesian framework (Hald et al., 2004). In 2008, a new methodological development was introduced (Pires and Hald, 2010), which enable the model to accommodate data from multiple years. This modification improved the robustness and accurateness of the results without compromising their comparability with estimates from previous years, and allows for the application of the model using data with less discriminatory power, e.g. with only serotyping as an epidemiological marker method (Pires and Hald, 2010).

The model routinely used in Denmark, as it is applied now, includes three dimensions: the *Salmonella* subtype, the animal-food sources (including imported food) and the year. It attributes sporadic cases of human salmonellosis to the animal-food sources, to outbreaks and to international travel each year. It is assumed that all cases that had been travelling abroad one week prior to onset of symptoms are travel-related. Because not all cases have travel information, human cases attributed to travel include cases that have reported to have travelled before onset of symptoms and estimated "extra-travelers", which constitute a proportion of the cases with unknown travel history that is estimated to have travelled based on the distribution of travelers and non-travelers for each subtype. A proportion of cases cannot be associated with any known source, and is gathered in a category named "Unknown source". This category includes cases caused by subtypes not found in any of the included sources or caused by isolates that were not subtyped. It may therefore include cases caused by sources not included in the model (e.g. game, seafood, etc).

Human cases caused by a subtype are attributed to animal sources based on the relative occurrence (represented by the prevalence) of this subtype in the animal sources included in the model. Two parameters are included in the model to take into consideration the ability of a subtype to cause disease and the ability of

[2] OJ L 268, 3.10.1998, p. 1–7. Decision No 2119/98/EC of the European Parliament and of the Council of 24 September 1998 setting up a network for the epidemiological surveillance and control of communicable diseases in the Community.

a source to convey the bacterial load to the human consumer. Those features enter the model as prior knowledge (flat priors), and this knowledge is updated by Bayesian inference (see 5.3.1) on the basis of the data available to inform the model: the number or reported cases caused by *Salmonella* subtypes and the prevalence of these subtypes in food/animal sources weighted by the amount of a food source available for consumption in the country.

4. HYPOTHESES AND OBJECTIVES

The overall aim was to explore ways to conduct source attribution studies in a global perspective, starting with the EU, where the data required for such studies were more readily available. This generated the following hypotheses:

- It is possible to develop an EU model based on the data available;
- It is possible to extrapolate results of the EU moel to countries with insufficient data using non-health indicators and expert elicitation;

To test the two hypotheses, the following specific objectives were set for parts I and II of the thesis:

Part I – The European Union model

4.1.1. To evaluate the quality and usefulness of the data available and do the necessary data management to include the maximum number of countries and animal sources in the EU model (Manuscript I).

4.1.2. To develop and run a Bayesian model based on microbial subtyping for attribution of human cases of salmonellosis in the EU (Manuscript II).

4.1.3. To compare the estimates obtained for Denmark with the results obtained for the same time period in the Danish model routinely applied, and propose improvements to both models. (Manuscript III).

Part II – An alternative approach for source attribution in countries with missing data

4.2.1. To propose and evaluate an alternative approach for source attribution based on expert elicitation, using non-health indicators as information to estimate results for countries, where the data on *Salmonella* required for the Bayesian model are not available.

Part I: The European Union Model

5. MATERIALS AND METHODS

Given the characteristics of the described methods, the data potentially available and the original proposal of developing a microbial subtyping-based model for this thesis, an adaptation of the three-dimensional Danish model was chosen as the best option for source attribution in the EU.

5.1. Data sources, handling and selection (Manuscript I)

5.1.1. The ideal dataset

The ideal dataset for a European model should have uniformly collected information, so results are comparable between countries. Considering the data requirements for the Danish three-dimensional source attribution model (Pires and Hald, 2010):

- number of reported salmonellosis cases in humans per year, with subtyping information (e.g. information on serovars, phage types of *S*. Typhimurium and *S*. Enteritidis and/or AMR profiles);
    - number of cases which had been travelling before onset of symptoms;
    - number of cases connected to outbreaks and, if available, the implicated food source;
- *Salmonella* prevalence in food-animals, broiler and layer flocks, pork, beef and imported foods, with subtyping information as for human data;
- amount of food sources available for consumption in the country per year;

It was proposed that the EU model should:

- have the year substituted as a dimension by country of attribution;
- take into consideration the country of origin of the foods being consumed in the country of attribution, by considering the amount of food items imported from different countries and the *Salmonella* prevalence in those food sources in those countries;

These features imply that the "perfect" dataset for the EU model would include the following information for all countries included:

- number of reported human salmonellosis cases, originating from a well-established surveillance system with national coverage in which cases are all confirmed by laboratory and subtyped to the furthest possible level;
- information on whether the person reported as a case had been travelling abroad one week prior to symptoms onset;
- number of cases connected to outbreaks and, if available, identified outbreak sources;
- prevalence of *Salmonella* subtypes, using the same subtyping method(s) as for human cases, in the maximum number of relevant animal reservoirs of the food chain;
- amount of an animal product available for consumption in each country;
- trade data: amount of food sources imported and exported within EU countries.

5.1.2. Data available in EU Member States

Data for the model were provided by EFSA, including datasets that were originally maintained by other institutions. Exceptions are duly noted in the text. The availability of data for sporadic human cases,

outbreaks and animal-food sources among countries are presented in Table 1. A detailed description of all data available, including data on consumption and trade of food sources, follows.

A list of 25 serovars was selected to be addressed, based on their occurrence and observed prevalence in humans and animals in the last years (EFSA, 2011a; EFSA, 2010a; EFSA, 2009a). The presence of those serovars in humans and the five animal sources in the countries studied is summarized in Table 2 of Manuscript I. For better visualization, only the 11 serovars most frequently found simultaneously in humans and animals are individually shown in the graphs and figures. Those are *S.* Enteritidis, *S.* Typhimurium, *S.* Kentucky, *S.* Virchow, *S.* Agona, *S.* Hadar, *S.* Derby, *S.* Newport and *S.* Stanley. *S.* Bovismorbificans was then included in the short list because of its emerging importance in humans in the last years (EFSA, 2010a, EFSA 2011b), totalizing 11 short-listed serovars.

Table 1. Availability of data from the different datasets by country.

Source	Data source[(a)]	Countries	Additional data sources
Laying hens	EUSR data 2008	AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GR, HU, IE, IT, LT, LU, LV, NL, NO, PL, PT, RO, SE, SI, SK, UK	
Cattle	EUSR data 2007-2009	AT, BE, BG, CH, CZ, DE, DK, EE, ES, FI, FR, GR, HU, IE, IT, LT, LU, LV, NL, NO, PL, PT, RO, SE, SI, SK, UK	FR: David, J (2009); LV: EUSR 2006
Pigs	BS 2006, lymph node	AT, BE, BG, CY, CZ, DE, DK, EE, ES, FI, FR, GR, HU, IE, IT, LT, LU, LV, NL, NO, PL, PT, SE, SI, SK, UK	
Broiler	BS 2008, carcasses	AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GR, HU, IE, IT, LT, LU, LV, MT, NL, NO, PL, PT, RO, SE, SI, SK, UK	GR: BS 2005/6
Turkey	BS 2006, Fattening turkeys	AT, BE, BG, CY, CZ, DE, DK, EE, ES, FI, FR, GR, HU, IE, IT, LT, LU, LV, NL, NO, PL, PT, SE, SI, SK, UK	EE: EUSR 2006; LU: EUSR 2008 LV: EUSR 2006;
Human cases	Foodborne outbreak data, 2007-2009	AT, BE, CH, CZ, DE, DK, EE, ES, FI, FR, HU, IE, LT, LV, NL, NO, PL, PT, RO, SE, SI, SK	
	TESSy case-based and aggregated data, 2007-2009[(b)]	AT, BE, BG, CY, CZ, DE, DK, EE, ES, FI, FR, GR, HU, IE, IT, LT, LU, LV, MT, NL, NO, PL, PT, RO, SE, SI, SK, UK	
	National monitoring and laboratory surveillance data 2007-2009[(c)]	PL, PT, NL, IT, DE	

(a) If data were missing from a specific source in a country, surrogate data sources used are indicated.
(b) Bulgaria reported human cases, but no serovar information was available.
(c) Obtained through direct contact with Member States.

5.1.2.1. *Salmonella* in humans

The effectiveness of a unified European surveillance system strongly depends on the quality of the national surveillance systems and the operational performance of the coordinating partners. Challenges to achieve such unification in an integrated and efficient way include differences in organization and effectiveness of existing national surveillance systems, which affects data comparability, as well as finding a

way to suitably disseminate results and improve overall data quality. In 2005, a strategy for infectious disease surveillance in Europe was developed, outlining the transition from the then project-based approach led by the Commission to a more coordinated, sustained approach managed by ECDC (ECDC, 2007).

In order to improve uniformity of reporting from Member States to the EU level, case definitions were developed by the European Commission[3] and put to use by MSs in 2003[4], with a revised version being adopted in 2008[5]. Such definitions were constructed in a way that enables reporting in the greatest extent possible, taking into account local differences in level of sensitivity and specificity, according to the different goals of information collection. Case definitions are used by MS for reporting to the ECDC and implemented in their national reporting systems, allowing more comparability of surveillance data within the EU.

These actions culminated in the creation of The European Surveillance System (TESSy), which is responsible for validation, cleaning, analysis and dissemination of data. All 27 EU MSs and the three EEA countries report their available data on 49 communicable diseases to the system. The list of diseases and the reporting steps are described in Decision No 2119/98/EC[6]. The EU-wide coverage, the cleaning and validation processes and the use of standardized case definitions makes TESSy the best available source for data on human diseases. Cases of salmonellosis are reported to TESSy and summarized in the European Union Summary Report on Trends and Sources of Zoonoses, Zoonotic Agents and Food-Borne Outbreaks (EUSR), published by the European Food Safety Authority (EFSA). According to the report (EFSA, 2011a), 108,618 lab-confirmed cases of human salmonellosis were reported by 27 MSs in 2009. Although those are the official numbers reported, the total for the Netherlands and Spain are extrapolations calculated based on sentinel surveillance systems with national coverage of 64% and 25%, respectively. Also, in specific years, Bulgaria, Lithuania, Poland and Romania only reported aggregated data, instead of case-based information.

In 2006, de Jong and Ekdahl estimated the risk of Swedish travelers being notified with salmonellosis when returning from 31 European countries, by comparing a Swedish dataset of travel-associated cases (within-Europe travel only) from 1997 to 2003 with a group of randomly selected Swedish residents with a history of recent travel to those same countries in the same period. This risk was then compared with official reporting data from such countries, and by using Norway as a reference value of 1, a set of country-specific multipliers were calculated to be used as correction factors for underreporting, representing the ratio between the true and reported cases, and thus allowing the calculation of more realistic incidence estimates (de Jong and Ekdahl, 2006). In 2012, Havelaar et al. (2012b) published an update of the underreporting factors (UF) by de Jong and Ekdahl; Swedish travelers' data from 2005 to 2009 were used to calculate the risk of being reported after returning from the Netherlands, which was multiplied with the incidence rate from a Dutch population-based study. *Salmonella* incidences were than calculated by multiplying the incidence rates with the population of each country, and the new UFs (Table 2) were obtained as the ratio between the "true" calculated incidences and the ones officially reported. The country-specific total of reported cases in 2007 to 2009 can be adjusted by applying those UFs, and the relative importance of each MS for the EU total can be

[3] OJ L 86 03.04.2002 p. 44-62. Commission Decision of 19 March 2002 laying down case definitions for reporting communicable diseases to the Community network under Decision No 2119/98/EC of the European Parliament and of the Council.
[4] OJ L 184 23.07.2003 p.3 5-39. Commission Decision of 17 July 2003 amending Decision No 2119/98/EC of the European Parliament and of the Council and Decision 2000/96/EC as regards communicable diseases listed in those decisions and amending Decision 2002/253/EC as regards the case definitions for communicable diseases.
[5] OJ L 159, 18.6.2008, p. 46–90. Commission Decision of 28 April 2008 amending Decision 2002/253/EC laying down case definitions for reporting communicable diseases to the Community network under Decision No 2119/98/EC of the European Parliament and of the Council.
[6] OJ L 268, 3.10.1998, p. 1–7. Decision No 2119/98/EC of the European Parliament and of the Council of 24 September 1998 setting up a network for the epidemiological surveillance and control of communicable diseases in the Community.

better assessed. This does not affect the decreasing tendency observed for the EU as a whole, as the same factors will be applied to all years, maintaining the proportions among them.

Table 2 shows the officially reported and the adjusted number of cases per country from 2007 to 2009, along with their relative contributions to the yearly EU totals and the UFs used for correction. The importance of the underreporting corrections is visible from the countries perspective, as much higher incidence rates can be calculated from the new totals, and from an European perspective when the contributions of each country to the total in the EU are compared. As an example, in 2009 Bulgaria appears in the official numbers as responsible for 1.1% of EU cases, which changes to 14.3% after the adjustment. This change could have a reflex on the attribution results, as the impact of the most important country-specific sources in the EU may change when the number of cases due to those sources is multiplied.

Table 2. Human cases of salmonellosis reported before and after adjusting for underreporting, UFs (with 95% Credibility Intervals) and relative contributions to the EU total, 2007-2009.

Country	UF (95% CI)[(a)]	2009				2008				2007			
		Reported		Adjusted		Reported		Adjusted		Reported		Adjusted	
		n	%	n	%	n	%	n	%	n	%	n	%
AT	11 (1.6 , 33.6)	2,775	2.6	30,525	0.5	2,312	1.7	25,432	0.3	3,386	2.2	37,246	0.5
BE	3.5 (0.3 , 12.5)	3,113	2.9	10,896	0.2	3,831	2.8	13,409	0.2	3,915	2.5	13,703	0.2
BG	718.5 (111.7 , 2140.5)	1,247	1.1	895,970	14.3	1,516	1.1	1,089,246	14.3	1,136	0.7	816,216	10.3
CY	173.2 (26.8 , 523.8)	134	0.1	23,209	0.4	169	0.1	29,271	0.4	158	0.1	27,366	0.3
CZ	28.9 (4.3 , 86.0)	10,480	9.6	302,872	4.8	10,707	8.0	309,432	4.1	17,655	11.5	510,230	6.4
DK	4.4 (0.7 , 13.1)	2,130	2.0	9,372	0.2	3,669	2.7	16,144	0.2	1,648	1.1	7,251	0.1
EE	16.9 (2.4 , 51.8)	261	0.2	4,411	0.1	647	0.5	10,934	0.1	428	0.3	7,233	0.1
FI	0.4 (0.0 , 1.2)	2,329	2.1	932	0.0	3,126	2.3	1,250	0.0	2,738	1.8	1,095	0.0
FR	26.9 (4.0 , 82.0)	7,153	6.6	192,416	3.1	7,186	5.3	193,303	2.5	5,313	3.5	142,920	1.8
DE	9.8 (1.5 , 29.3)	31,395	28.9	307,671	4.9	42,885	31.9	420,273	5.5	55,399	36.0	542,910	6.8
GR	1228.5 (188.5 , 3668.2)	403	0.4	495,086	7.9	792	0.6	972,972	12.8	706	0.5	867,321	10.9
HU	66.8 (10.2 , 199.1)	5,873	5.4	392,316	6.3	6,637	4.9	443,352	5.8	6,578	4.3	439,410	5.5
IE	5.4 (0.0 , 27.2)	335	0.3	1,809	0.0	447	0.3	2,414	0.0	440	0.3	2,376	0.0
IT	71.7 (10.7 , 214.0)	4,156	3.8	297,985	4.8	6,662	5.0	477,665	6.3	6,731	4.4	482,613	6.1
LV	43.3 (6.6 , 134.9)	798	0.7	34,553	0.6	1,229	0.9	53,216	0.7	619	0.4	26,803	0.3
LT	59.1 (8.7 , 182.1)	2,063	1.9	121,923	2.0	3,308	2.5	195,503	2.6	2,270	1.5	134,157	1.7
LU	4.5 (0.0 , 21.4)	162	0.1	729	0.0	153	0.1	689	0.0	163	0.1	734	0.0
NL	26.3 (3.6 , 84.8)	1,205	1.1	31,692	0.5	1,627	1.2	42,790	0.6	1,224	0.8	32,191	0.4
PL	114.1 (17.2 , 338.2)	8,521	7.8	972,246	15.6	9,148	6.8	1,043,787	13.7	11,155	7.3	1,272,786	16.0
PT	2082.9 (318.3 , 6266.9)	220	0.2	458,238	7.3	332	0.2	691,523	9.1	438	0.3	912,310	11.5
RO	349.9 (48.0 , 1127.8)	1,105	1.0	386,640	6.2	624	0.5	218,338	2.9	620	0.4	216,938	2.7
SK	53.2 (7.6 , 165.4)	4,182	3.9	222,482	3.6	6,849	5.1	364,367	4.8	8,367	5.4	445,124	5.6
SE	40.3 (4.9 , 133.2)	616	0.6	24,825	0.4	1,033	0.8	41,630	0.5	1,336	0.9	53,841	0.7
ES	214.2 (32.7 , 638.9)	4,304	4.0	921,917	14.8	3,833	2.8	821,029	10.8	3,842	2.5	822,956	10.4
SE	0.5 (0.1 , 1.6)	3,054	2.8	1,527	0.0	4,185	3.1	2,093	0.0	3,930	2.6	1,965	0.0
UK	7.3 (1.1 , 22.6)	10,479	9.6	76,497	1.2	11,511	8.6	84,030	1.1	13,557	8.8	98,966	1.2
EU-27	57.5 (8.8 , 171.4)	108,618	100	6,245,535	100	134,579	100.0	7,738,293	100	153,837	100.0	8,845,628	100

(a) Havelaaer at al., 2012. (b) EFSA 2012a

5.1.2.1.1. Human data selection and handling

Data on the number and serovar distribution of human cases reported to TESSy from 2007 to 2009, as provided by ECDC through EFSA, were extracted on 6th of July 2010. The total number of reported cases includes sporadic, travel-related and outbreak-related infections.

Travel data

Travel information was reported as "imported", "not-imported" or "unknown location of origin", and the amount of information actually provided varied in frequency and quality. The proportion of travelers varied greatly among MSs, and for some countries, such as Sweden and Finland, travel-related infections were the majority of all salmonellosis cases, while in France, Romania and Slovenia, 100% of cases had unknown travel information. Full travel information was provided by Austria, Belgium, the Czech Republic, Estonia, Spain, Hungary, the Netherlands and Slovakia (Table 4).

Outbreak data

For outbreaks of foodborne salmonellosis, the same datasets used for the EUSRs 2007-2009 were provided by EFSA. For data management and modeling purposes, it was assumed that countries which reported sporadic cases but no outbreak cases did not have any foodborne *Salmonella* outbreaks in the period.

Dealing with missing information

MSs for which the level of serovar detailing was insufficient for source attribution were requested to provide additional data, if available. Such national datasets with more detailed serovar information were provided by Poland and Portugal.

From the TESSy data, out of 392,485 cases reported in the EU, 35,643 (9.1%) had incomplete or missing serovar information. This rate varied within countries, from zero in Portugal to 84% in Romania. Types of incompleteness detected varied as to how far the identification reached (Figure 3), and can be summarized as:

a) classification up to genus or species level, such as *Salmonella* spp, or *Salmonella enterica* (0.02% of EU cases);

b) classification up to subspecies level, such as *Salmonella enterica enterica* or *Salmonella ente*rica Subspecies I (0.2%);

c) classification using groups based on the O-antigen both by the old nomenclature (groups B, C1-C2 or E4) or the new one (serogroups O:4, O:7 or O:33) (2.3%);

d) aggregated data, where the main serovars were specififed, and the remaining were grouped as "Others" (3.9%);

e) cases where the serovar field was simply blank or filled with "unknown" (2.7%).

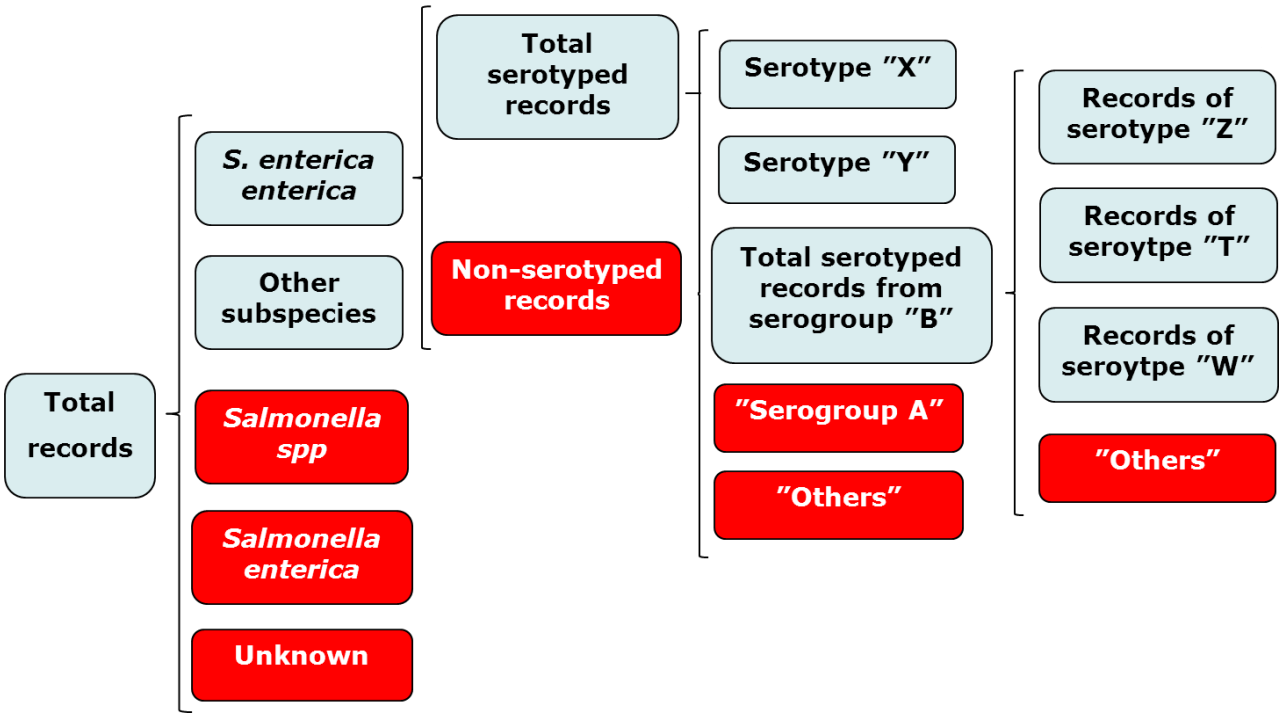


Figure 3. Observed "levels" of serovar identificationfound in human and animal datasets. Red boxes show situations which required serovar reassignment.

Reported human isolates that were not classified up to serovar level or data reported in aggregated form were reassigned to specific serovars according to proportions observed in previous studies, in the same dataset or in other references, depending on the availability of data in each case. So, "*Salmonella* spp.", "*Salmonella* spp., unspecified", "*Salmonella* untyped", "*Salmonella* not typed", "*Salmonella enterica enterica*" and "*Salmonella* Subsp. I" were reassigned to all serovars observed in the country. (e.g.: If *S.* Enteritidis accounted for 60% of all serotyped isolates from human cases in a country, and 10 isolates in the same country received one of the denominations mentioned, six of them were reassigned to *S.* Enteritidis, and so on for other observed serovars). Isolates identified up to subspecies level should likewise be reassigned to all serovars in the country, but with proportions calculated using only isolates of *S. enterica enterica* as total.

Isolates classified as serogroups (e.g. C or D) or subgroups (e.g. C1 or D1) were distributed among serovars pertaining to those groups, in accordance with the Kauffman-White-Le Minor Scheme 9th edition (WHOCC-Salm, 2007) (e.g., if *S.* Typhimurium accounted for 40% of all serotyped human isolates in the country, but for 80% of isolates from serovars belonging to Group B, and 10 isolates were identified as "*Salmonella* Group B", eight of those were reassigned to *S.* Typhimurium, and so on for other serovars of the same group).

Isolates classified as "Others", "Other" or "Other serovars" were assumed to belong to serovars not described in the current dataset, but nonetheless present in the country. In this case, the main reference dataset used to obtain the proportions for the reassignment was the WHO GFN CDB (http://thor.dfvf.dk/gss), which contains the 15 most commonly identified *Salmonella* serovars among human and non-human sources in 84 countries (e.g.: suppose that in the original TESSy data, a country reported 30 isolates: 10 *S.* Enteritidis, 10 *S.* Typhimurium and 10 "Others". The GFN CDB showed 80% *S.* Enteritids, 10% *S.*

Typhimurium, 7% *S.* Infantis and 3% *S.* Hadar for this country, so, according to this reference, *S.* Infantis and *S.* Hadar correspond to 70% and 30% of the non-described serovars. The 10 records were then redistributed as seven *S.* Infantis and three *S.* Hadar, as it was assumed that no *S.* Typhimurium or *S.* Enteritidis isolates were included in the group of "others").

One of the defining antigenic characteristics of *S.* Typhimurium is that it possesses two phases of H-antigens: "i" and "1,2", which is why the antigenic formula for this serovar is written as "1,4,[5],12:i:1,2" (WHOCC-Salm 2007). However, variants that lack either the first or the second phase H antigen have been described, and reported by some countries as "1,4,[5],12:i:-" or "1,4,[5],12:-:1,2". Those variants are referred to as "*S.* Typhimurium-like strains" or, in the cases cited, "1,4,[5],12:i:-", "Monophasic *S.* Typhimurium". For our purposes, isolates identified by those formulas in the datasets were reassigned to *S.* Typhimurium, which is supported by an EFSA BIOHAZ Panel assessment (EFSA, 2010b), as not all countries made that differentiation, and it is assumed that those variants were already reported as *S.* Typhimurium by some of them. The aphasic antigenic formula or "1,4,[5],12:-:-" was not reassigned, as it could belong to several serovars in group O:4.

Table 3 shows the number and percentage of total reported and reassigned records by type of inconsistency. Total reported cases may differ from totals reported in the EUSR, as the datasets were extracted in different dates. The totals shown, however, have been validated and accepted by the MSs for publication in Pires et al. (2012).

Travel information was often found incomplete. All records with missing or unknown travel information from countries which presented serovar detailing of sporadic cases were considered domestically acquired in the reporting country. In previous models, such as the Danish source account model, cases with unknown or blank travel history were reassigned to travel or domestic cases, based on the proportions among those two categories. In this dataset, several countries reported 100% of cases without information, and so it was not possible to proportionally reassign them. The assumption of non-travel in case of no information was kept for countries which had some travel history, to keep coherence in the method across all countries. Furthermore, it was considered that assuming all cases without travel information were travel-related would be a less sustainable assumption than that they were all domestic. The number of cases with missing travel information and the final number included in each country are shown in Table 4.

Outbreak-related cases for which a serovar was not fully identified were reassigned using the proportions observed in the same dataset, as some serovars may be more prone to generate outbreaks than others, and thus the proportions observed in reported sporadic cases may not apply. Bulgaria, Cyprus, Greece, Italy, Luxembourg, Malta and the United Kingdom did not report any cases. As all of those countries, except for Bulgaria, had properly reported sporadic cases during the three years under study, it was assumed that they had no outbreak cases to be reported in the period. France needed to have nearly 47% of outbreak cases reassigned, as the isolates were reported as "*Salmonella* spp.", the same happening with 39% of isolates in Latvia. The number and percentage of reassigned outbreak records are shown in Table 5.

Table 3. Number and percentage of reassigned records in humans.

Country	Incomplete identification						Aggregated data[(d)]		Unknown[(e)]		Total		
	Species/genus[(a)]		Subspecies[(b)]		Serogroup[(c)]						Reported	Reassigned	
	n	%	n	%	n	%	n	%	n	%		n	%
AT			2	0.02	132	1.56	287	3.38	362	4.27	8,487	783	9.23
BE							172	1.55			11,066	172	1.55
BG	-	-	-	-	-	-	-	-	-	-	3,899	-	-
CY	2	0.42			9	1.91			101	21.44	471	112	23.78
CZ									586	1.51	38,842	586	1.51
DE			462	0.36	8,057	6.33	5,782	4.54	1,628	1.28	127,330	15,929	12.51
DK			2	0.03	3	0.04	25	0.33	342	4.56	7,497	372	4.96
EE					25	1.86	28	2.09			1,341	53	3.95
ES							2,504	20.81	2,091	17.38	12,033	4,595	38.19
FI	19	0.23	3	0.04	23	0.28	6	0.07	22	0.27	8,228	73	0.89
FR							2,185	10.75			20,319	2,185	10.75
GR					104	5.40	3	0.16	1,309	67.93	1,927	1,416	73.48
HU			57	0.30	191	1.00	908	4.76	2	0.01	19,091	1,158	6.07
IE	1	0.08					11	0.87	68	5.38	1,264	83	6.57
IT	25	0.24			6	0.06			1,080	10.58	10,205	1,111	10.89
LT					56	0.73	156	2.04	191	2.50	7,643	403	5.27
LU									63	13.15	479	63	13.15
LV							53	1.99	608	22.81	2,665	661	24.80
MT	20	5.39							40	10.78	371	60	16.17
NL			210	5.04			84	2.02			4,168	294	7.05
PL							1204	3.89			30,963	1,204	3.89
PT											1,513	0	0.00
RO							1,218	51.81	766	32.58	2,351	1,984	84.39
SE			68	0.60			411	3.65	307	2.73	11,265	786	6.98
SI					63	2.10					3,002	63	2.10
SK	3	0.02			154	0.79	84	0.43	87	0.45	19,399	328	1.69
UK	7	0.02			149	0.41	4	0.01	1,009	2.75	36,666	1,169	3.19
EU total	77	0.02	804	0.20	8,975	2.29	15,125	3.85	10,662	2.72	392,485	35,643	9.08
CH	-	-	-	-	-	-	-	-	-	-	-	-	-
NO							21	0.44	10	0.21	4825	31	0.64
Total	77	0.02	804	0.20	8,975	2.26	15,146	3.81	10,672	2.69	397,310	35,674	8.98

(a) *Salmonella* spp, *Salmonella enterica*, *Salmonella* not typed, *Salmonella* untyped
(b) *Salmonella enterica enterica*, Subspecies I
(c) B, C, D, E, D1, C1, C2-C3, D1, E1
(d) ”Others”, ”Other serovars”, ”Unknown”

Table 4. Number of cases reported in the original datasets as travel-related, domestic or unknown and the total used in the model, assuming that any case not specifically mentioned as travel-related was domestic.

Country	Reported			Total used	
	Travel	Domestic	Unknown	Travel	Domestic
AT	988	7,499	0	988	7,499
BE	0	11,066	0	0	11,066
BG	-	-	-	-	-
CY	18	428	25	18	453
CZ	657	38,185	0	657	38,185
DE	6,683	114,362	6,285	6,683	120,647
DK	1,366	2645	3,486	1,366	6,131
EE	95	1246	0	95	1,246
ES	0	12,033	0	0	12,033
FI	6,845	1059	324	6,845	1,383
FR	0	0	20,319	0	20,319
GR	45	1763	119	45	1,882
HU	29	19,062	0	29	19,062
IE	384	343	537	384	880
IT	132	692	9,381	132	10,073
LT	21	0	7,622	21	7,622
LU	46	431	2	46	433
LV	32	1,817	816	32	2,633
MT	4	365	2	4	367
NL	497	3,671	0	497	3,671
PL	16	0	30,947	16	30,947
PT	5	0	1,508	5	1,508
RO	0	0	2,351	0	2,351
SE	8,752	2,207	306	8,752	2,513
SI	0	0	3,002	0	3,002
SK	146	19,253	0	146	19,253
UK	8,921	8,084	19,661	8,921	27,745
EU total	35,682	246,211	106,693	35,682	356,803

Table 5. Number and percentage of reassigned records in foodborne *Salmonella* outbreaks.

Country	Reported	Incomplete identification				Total		
		Species/genus[(a)]		Serogroup[(b)]		Reported	Reassigned	
		n	%	n	%		n	%
AT	Yes					421	0	0.00
BE	Yes					91	0	0.00
BG	No					-	-	-
CY	No					0	0	0.00
CZ	Yes					337	0	0.00
DE	Yes	13	0.55			2,383	13	0.55
DK	Yes					2,224	0	0.00
EE	Yes					157	0	0.00
ES	Yes					469	0	0.00
FI	Yes					189	0	0.00
FR	Yes	1218	46.68			2,609	1,218	46.68
GR	No					0	0	0.00
HU	Yes	86	4.48			1,921	86	4.48
IE	Yes					67	0	0.00
IT	No					0	0	0.00
LT	Yes					371	0	0.00
LU	No					0	0	0.00
LV	Yes	201	39.26			512	201	39.26
MT	No					0	0	0.00
NL	Yes	12	1.71	26	3.71	700	38	5.43
PL	Yes			29	0.55	5,310	29	0.55
PT	Yes					90	0	0.00
RO	Yes	26	5.95			437	26	5.95
SE	Yes	8	2.94			272	8	2.94
SI	Yes					692	0	0.00
SK	Yes					583	0	0.00
UK	No					0	0	0.00
EU total	-	1,564	7.89	55	0.28	19,835	1,619	8.16
CH	Yes					6	0	0.00
NO	Yes					95	0	0.00
Total	-	1,564	7.85	55	0.28	19,936	1,619	8.12

(a) *Salmonella enterica enterica*, Subspecies I
(b) B, C, D, E, D1, C1, C2-C3, D1, E1

Serovar information

S. Enteritidis and *S.* Typhimurium are the serovars most frequently observed in humans in Europe (EFSA, 2011a). Table 6 shows that the accompanying list of serovars may vary from year to year (EFSA, 2012a; EFSA, 2010a), depending on the occurrence of outbreaks or changes in the surveillance and monitoring of food and animal sources. However, a smaller list has been constantly observed in the last five years, besides *S.* Enteritidis and *S.* Typhimurium, namely *S.* Infantis, *S.* Newport, *S.* Kentucky, *S.* Virchow, *S.* Derby and *S.* Agona. The decreasing trend in *Salmonella* cases has a visible reflection in the proportion of cases due to *S.* Enteritidis; as most of the control measures started at EU-level during the decade of 2000 have been applied to eggs and layers[7], there are proportionally fewer cases of *S.* Enteritidis every year, resulting in a relative increase on the reported proportions of other serovars (Table 6).

The most common serovars observed in outbreaks were *S.* Enteritidis and *S.* Typhimurium. As expected, outbreaks may happen associated with serovars not normally found in the country. That is particularly true in countries with a small number of sporadic cases and a good level of control of *Salmonella* in domestic products, which is exemplified in Figure 4 when comparing the serovar profile of sporadic and outbreak cases in Finland, Sweden and Norway.

As phage type reporting is not mandatory by EU regulations, this information is not complete or harmonized, being only available for 62.0% of isolates from the United Kingdom, 61.9% from Denmark and 26.6% from Austria. The remaining MSs phage typed between zero and 15% of total reported isolates. Thus, phage type information in humans was considered not useful, as the limited availability of data renders this subtyping level impossible to use in the model.

Table 6. Distribution of salmonellosis cases in humans (%) in the eleven selected serovars (TESSy), 2007-2009.

Serovar	**2009[(a)]**	**2008[(b)]**	**2007[(b)]**
S. Enteritidis	52.3	58.0	64.5
S. Typhimurium	23.3	21.9	16.5
S. Infantis	1.6	1.1	1.0
S. Newport	0.7	0.7	0.6
S. Kentucky	0.5	0.4	0.3
S. Virchow	0.7	0.7	0.8
S. Derby	0.7	0.5	0.4
S. Agona	-	0.5	0.3
S. Hadar	0.5	-	-
S. Bovismorbificans	0.4	0.4	-
S. Stanley	-	0.4	0.5
Other[(dc)]	18.8	15.3	14.7

(a) EFSA 2012a; (b) EFSA 2010a; (c) For each year, this category includes serovars not among the top-10, even if they are present in other years

[7] OJ L 325, 12.12.2003, p. 1. Regulation as last amended by Commission Regulation (EC) No 1237/2007 (OJ L 208, 24.10.2007, p. 5)

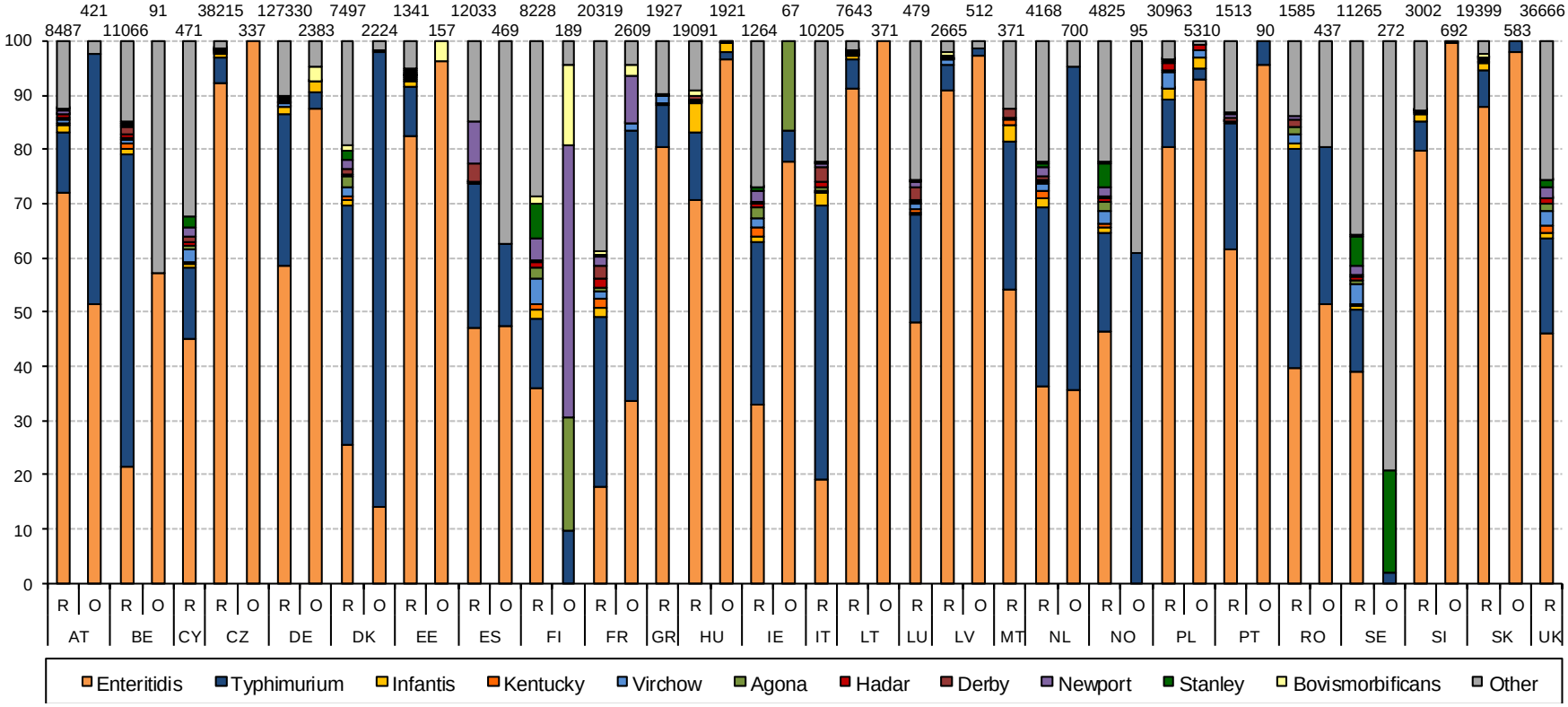


Figure 4. Relative proportions of the most frequent serovars in total reported (R) and outbreak (O) cases in humans in the EU and Norway, 2007-2009. The totals for each country in the datasets are shown at the top of the bar.

5.1.2.2. *Salmonella* in animal reservoirs of the food chain

The European Parliament Regulation 2160/2003/EC[8] on the control of *Salmonella* and other specified foodborne zoonotic agents has as main objective the reduction of *Salmonella* in animal populations at farm level. In order to provide the scientific basis for setting prevalence targets in commercial and breeding farms, EU-wide studies on the baseline prevalence of *Salmonella* were conducted focusing on laying hens (2004-2005), broiler flocks (2005-2006), slaughter pigs (2006-2007), fattening and breeding turkeys (2006-2007), broiler batches at slaughter (carcasses) (2008) and breeder pigs (2008). The studies took place during a four-year period and varied in MS participation due to the addition of new members to the EU in 2004 and 2007 and the occasional participation of non-MS, such as Norway and Switzerland. However, they still constitute the most uniformly collected and analyzed data on *Salmonella* at the EU level, thus allowing valid comparisons among MSs. No harmonized data from *Salmonella* in cattle or other sources were available.

Besides supervising the Baseline Studies (BS), EFSA is responsible for examining the data on zoonoses, antimicrobial resistance and foodborne outbreaks collected from MSs in accordance with Directive 2003/99/EC[9] and publishing those results annually in the EUSR (EFSA, 2007a). Data on the occurrence of zoonotic agents in animals, foodstuffs and animal feed are reported directly by MSs to EFSA, but not all member states report all categories, and serovar information is frequently reported in aggregated form (EFSA, 2012a). The quality and comparability of the data have been improved in recent years, since, as targets are being set for the reduction of certain *Salmonella* serovars in different poultry populations[10], the monitoring of layers (EFSA, 2009b), broilers (EFSA, 2011b) and, more recently, turkeys (EFSA, 2012b) have been harmonized at EU level. However, this harmonization occurred after the period comprised by this thesis, with the exception of the monitoring of layers, and so at that moment, the most uniform data source for *Salmonella* in the other sources in the EU were the EFSA BS.

5.1.2.2.1. Animal data selection and handling

Data from the EU BS on the prevalence of *Salmonella* in broiler carcasses (EFSA, 2010c), slaughter pigs (EFSA, 2008a) and fattening turkeys (EFSA, 2008b) were used. These datasets were considered the most representative of the given reservoir, since no harmonized EU monitoring in pigs and turkeys is currently in place, and the broiler carcass study was considered to provide sufficiently recent data with a better detailing of the serovar distribution, when compared to the existing EU monitoring data.

In order to use the most recent data possible, the laying hens BS (EFSA, 2007c) was not used. The study was conducted between 2004 and 2005, and it is expected that the implementation of the Commission Regulation (EC) No 1168/2006[11] for harmonizing the surveillance of laying hens flocks of *Gallus gallus* in the EU has resulted in significant changes in the *Salmonella* serovar prevalences in this reservoir in many MSs. Instead, data for laying hens were obtained from the EUSR 2008 (EFSA, 2010a), which was the first year of EU-harmonised reporting for this reservoir. Selection of data by country was performed according to the recommendations found in EFSA (2010d). Cattle data were retrieved from the EUSR 2007, 2008 and 2009 (EFSA, 2009a; EFSA, 2010a; EFSA, 2011a), with 2009 data being preferred to the other years.

[8] EUT L 325 af 12.12.2003, s. 1–15.
[9] OJ L 325, 12.12.2003, p. 31–40, amending Council Decision 90/424/EEC and repealing Council Directive 92/117/EEC
[10] OJ L 325, 12.12.2003, p. 1. Regulation as last amended by Commission Regulation (EC) No 1237/2007 (OJ L 208, 24.10.2007, p. 5)
[11] OJ L 211, 1.8.2006, p. 4–8, implementing Regulation (EC) No 2160/2003 and amending Regulation (EC) No 1003/2005.

The epidemiological unit used for data description varied with the study, being carcasses, flocks, herds or individual samples, depending on the data used. Study design, sampling schemes, sample units and data collection methods can be found in the respective reports (EFSA, 2007b; EFSA, 2008a; EFSA, 2008b; EFSA, 2010c). Prevalence of *Salmonella* serovars in the different food-animal sources were calculated to describe the available data by dividing the number of positive units (samples or herds/flocks) by the number of tested units and multiplying the result by one hundred. The values obtained may differ from the original prevalences in the BS; this happened because those were calculated as EU-weighted prevalences, but only for the main serovars, and serovar-specific prevalences are needed for the modeling. However, as cross-country prevalence comparisons were not intended here, and given the coverage achieved by the sampling in the studies, the non-weighted values are valid as local prevalences for each country. To estimate relative frequencies, the numerator was the number of units positive for a specific serovar and the denominator was the number of total positive units.

Data were available from 28 countries in different combinations of animal data sources. Highest positivity at EU level was observed for turkeys (20.7%), followed by pigs (13.9%), broilers (13.1%), laying hens (5.9%) and cattle (4.5%) (Table 7). Given the non-uniformity of the data collection for cattle, interpretation of these estimates should be made with care. Belgium and the United Kingdom only reported positive samples for cattle, resulting in 100% positivity. Small samples were also observed for broilers in Luxembourg, laying hens in Lithuania and Luxembourg, and turkeys in Estonia, Luxembourg and Latvia (Table 16). These small samples showed a very low or, in most cases, zero positivity, but care should be taken when looking at sample sizes that could be representative of a small national production, but which could also be an imprecise reflection of the animal population in those countries.

Dealing with missing information

Greece did not take part in the 2008 broiler carcasses BS, so serovar information in this country was supplied with data from the broiler flocks BS, conducted between 2005 and 2006 (EFSA, 2007b). For slaughter pigs, the results of the lymph node sampling were available for most MSs, except for Malta and Romania (EFSA, 2008a). BS data from fattening turkeys were used, with the exception of Estonia, Latvia, Luxembourg and Romania, which were not part of the study. Consequently, data on *Salmonella* serovars in turkeys from these countries were provided by EUSR data from 2006 and 2008 (EFSA, 2007a; EFSA, 2010a), except for Romania, from where no data were available.

In the laying hens data, in addition to units with non-identified or partially identified serovars, many countries only reported a reduced list of serovars and a group of "Others". Sometimes this reduced list would comprise only *S.* Enteritidis and *S.* Typhimurium, as those are the two serovars for which specific reporting is mandatory by Directive 2003/99/EC. For that reason, the proportions used for re-allocation of units were the ones found in the EFSA Laying Hens Baseline Study (EFSA, 2007c). For BS data, no reference for reassigning of serogroups or incomplete serovar identification was available, so units were redistributed according to the proportions found among serovars in the same dataset.

For cattle, no data from Cyprus or Malta were identified, and serovar information for France was supplemented with data from David (2009).

The same criteria as for humans were used for reassigning non-identified or partially identified serovars. The detailed amount and percentage of reassigned records among the total positives in the BSs or EUSR are shown in Table 8.

Table 7: Number of sampling units submitted and positivity percentages in animal reservoirs in the EU and Norway.

Country	Broiler carcasses[(a)]			Pigs – lymph nodes			Laying hen flocks			Turkeys – fattening flocks			Cattle [(b)]		
	Submitted	Positives		Submitted	Positives		Submitted	Positives		Submitted	Positives		Submitted	Positives	
		n	%		n	%		N	%		N	%		N	%
AT	408	10	2.5	617	13	2.1	1,966	49	2.5	1,010	141	14.0	3,037	12	0.4
BE	380	77	20.3	601	78	13.0	649	76	11.7	370	40	10.8	81	81	100.0
BG	316	85	26.9	176	35	19.9	119	0	0.0	85	0	0.0	477	3	0.6
CY	357	38	10.7	359	47	13.1	40	5	12.5	70	28	40.0	-	-	-
CZ	422	23	5.5	654	38	5.8	449	40	8.9	970	192	19.8	696	24	3.4
DE	432	76	17.6	2,567	325	12.7	6304	220	3.5	1,475	108	7.3	4,053	163	4.0
DK	396	0	0.0	998	80	8.0	508	3	0.6	294	1	0.3	7,915	9	0.1
EE	102	0	0.0	420	27	6.4	52	4	7.7	2	0	0.0	1,550	10	0.6
ES	389	58	14.9	2,621	806	30.7	845	376	44.5	1,910	747	39.1	258	29	11.2
FI	369	0	0.0	419	0	0.0	950	1	0.1	675	0	0.0	3,415	7	0.2
FR	422	32	7.6	1,163	215	18.5	3067	187	6.1	1,630	157	9.6	-	-	2.4
GR	1,215	180	14.8	345	73	21.2	112	35	31.3	220	16	7.3	56	1	1.8
HU	321	275	85.7	656	75	11.6	866	101	11.7	1,465	915	62.5	178	31	17.4
IE	394	39	9.9	422	65	15.4	204	2	0.98	1,295	294	22.7	10,121	430	4.2
IT	393	66	16.8	709	116	16.4	821	171	20.8	1,370	277	20.2	1,797	17	0.9
LT	374	26	6.9	461	8	1.7	13	0	0.0	315	14	4.4	172	2	1.2
LU	13	0	0.0	313	50	16.0	7	1	14.3	1	0	0.0	83	7	8.4
LV	122	6	4.9	392	21	5.4	69	14	20.3	1	0	0.0	25	0	0.0
MT	367	77	21,0	-	-	-	-	-	-	-	-	-	-	-	-
NL	429	43	10.0	1,087	92	8.5	2346	62	2.6	860	77	9.0	330	18	5.5
PL	419	107	25.5	1,176	75	6.4	1533	192	12.5	1,610	285	17.7	130	0	0.0
PT	421	47	11.2	658	156	23.7	227	83	36.56	525	26	5.0	56	0	0.0
RO	357	17	4.8	-	-	-	-	-	-	-	-	-	521	3	0.6
SE	410	1	0.2	394	6	1.5	724	5	0.7	70	0	0.0	3,728	60	1.6
SI	413	7	1.7	431	27	6.3	172	18	10.5	655	100	15.3	386	1	0.3
SK	422	91	21.6	385	30	7.8	138	10	7.2	125	15	12.0	95	0	0.0
UK	401	14	3.5	639	139	21.8	5523	67	1.2	1,570	401	25.5	895	895	100.0
EU Total	9,249	1,215	13.1	18,663	2,596	13.9	27,704	1630	5.9	18,514	3,834	20.7	40,055	1,803	4.5
NO	396	0	0.0	408	1	0.2	1080	0	0.0	360	0	0.0	2,589	1	0.0
Total	10,035	1,225	12.2	19,072	2,598	13.6	28,784	1630	5.7	18,849	3,834	20.3	42,644	1,804	4.2

(a) In the specific case of Greece, broiler flocks. (b) In the specific case of Denmark, carcass samples collected at the slaughterhouse.

Table 8. Number and percentage of records reassigned to serovars in animal reservoirs.

	Country	Incomplete identification						Aggregated [(d)]		Total		
		Species/genus[(a)]		Subspecies[(b)]		Serogroup[(c)]				Positives	Reassigned	
		n	%	n	%	n	%	n	%		n	%
Broilers	BE	15	19.48							77	15	19.48
	IT	13	19.70							66	13	19.70
	LT	15	57.69							26	15	57.69
	MT	10	12.99							77	10	12.99
	NL	1	2.33							43	1	2.33
Pigs	BG			4	11.43					35	4	11.43
	CY	5	10.64	3	6.38	1	2.13			47	9	19.15
	DE	5	1.54			64	19.69			325	69	21.23
	EE			4	14.81					27	4	14.81
	ES	62	7.69							806	62	7.69
	FR	5	2.33							215	5	2.33
	GR	3	4.11	8	10.96					73	11	15.07
	IE	1	1.54							65	1	1.54
	IT	41	35.34	6	5.17					116	47	40.52
	LV	2	9.52							21	2	9.52
	NL	2	2.17	2	2.17					92	4	4.35
	SI	4	14.81							27	4	14.81
Turkeys	CY					5	17.86			28	5	17.86
	DE					11	10.19			108	11	10.19
	DK	1	100.00							1	1	100.00
	HU	1	0.11	2	0.22					915	3	0.33
	IT			8	2.89					277	8	2.89
	SI					1	1.00			100	1	1.00
Layers	AT	2	4.08							49	2	4.08
	BE	3	3.95			3	3.95			76	6	7.89
	CY					1	20.00			5	1	20.00
	DE	13	5.91					23	10.45	220	36	16.36
	ES	186	49.47							376	186	49.47
	FR	20	10.70					6	3.21	187	26	13.90
	HU							26	25.74	101	26	25.74
	IT							115	67.25	171	115	67.25
	PL							29	15.10	192	29	15.10
	PT							9	10.84	83	9	10.84
	UK							16	23.88	67	16	23.88
Cattle	BE	3	3.70			4	4.94			81	7	8.64
	DE	4	2.45					36	22.09	163	40	24.54
	DK	4	44.44							9	4	44.44
	ES	13	44.83							29	13	44.83
	HU	25	80.65							31	25	80.65
	IT	4	23.53							17	4	23.53
	LU	1	14.29							7	1	14.29
	NL	1	5.56							18	1	5.56
	SE	6	10.00							60	6	10.00
	UK	824	92.07							895	824	92.07

(a) *Salmonella* spp, *Salmonella enterica*, *Salmonella* not typed, *Salmonella* untyped
(b) *Salmonella enterica enterica*, Subspecies I
(c) B, C, D, E, D1, C1, C2-C3, D1, E1
(d) ”Others”, ”Other serovars”

Serovar information

The main serovars vary from source to source (Table 9). As an example, *S.* Orion was only found in turkeys, and the importance of *S.* Infantis and *S.* Enteritidis in *Gallus gallus, S.* Typhimurium and *S.* Derby in pigs and *S.* Bredeney and *S.* Saintpaul in turkeys is much larger than in other sources. This is an important feature for source attribution, as it helps the tracking of animal sources when cases of those serovars are reported in humans.

Table 9. Relative proportions of the top-10 serovars found in broiler carcasses, pig lymph nodes, turkey flocks and laying hen flocks in the EFSA Baseline Studies.

Serovar[(a)]	Broilers[(b)]	Pigs[(c)]	Turkeys[(d)]	Layers[(e)]
S. Infantis	**29.2**	**1.9**	**6.6**	**11.5**
S. Enteritidis	**13.6**	**4.9**	**5.1**	**59.9**
S. Kentucky	**6.2**	0.0	0.1	0.0
S. Typhimurium	**4.4**	**44.9**	**7.9**	**8.3**
S. Bredeney	**4.3**	**2.0**	**17.2**	1.0
S. Virchow	**4.1**	0.3	1.0	**2.7**
S. Hadar	**3.8**	0.3	**14.0**	**3.4**
S. Paratyphi var. Java	**3.8**	0.1	0.2	0.1
S. Agona	**3.0**	**1.1**	2.9	**2.2**
S. Indiana	**2.9**	0.1	3.0	0.3
S. Derby	0.8	**14.6**	**11.3**	0.0
S. Rissen	0.0	**5.8**	0.0	0.5
S. Anatum	0.7	**2.4**	0.4	0.7
S. London	0.0	**1.3**	2.9	0.0
S. Brandenburg	0.2	**1.2**	0.0	0.9
S. Saintpaul	0.2	0.1	**10.3**	0.0
S. Kottbus	0.7	0.3	**8.3**	0.0
S. Orion	0.0	0.0	**6.1**	0.0
S. Blockley	1.8	0.1	**3.7**	0.0
S. Mbandaka	2.4	0.3	0.8	**6.6**
S. Livingstone	1.0	0.4	0.0	**3.4**
S. Ohio	0.9	0.3	0.0	**2.4**
S. Braenderup	0.2	0.2	0.1	**2.0**

(a) Combined list of the top ten serovars in all BS. Top-ten serovars for each source have values in bold.
(b) EFSA (2010c). Participating countries (29): AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GR, HU, IE, IT, LT, LU, LV, MT, NL, NO, PL, PT, RO, SE, SI, SK, UK;
(c) EFSA (2008a). Participating countries (26): AT, BE, BG, CY, CZ, DE, DK, EE, ES, FI, FR, GR, HU, IE, IT, LT, LU, LV, NL, NO, PL, PT, SE, SI, SK, UK.
(d) EFSA (2008b). Participating countries (26): AT, BE, BG, CY, CZ, DE, DK, EE, ES, FI, FR, GR, HU, IE, IT, LT, LU, LV, NL, NO, PL, PT, SE, SI, SK, UK.
(e) EFSA 2010a. Participating countries (28): AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GR, HU, IE, IT, LT, LU, LV, NL, NO, PL, PT, RO, SE, SI, SK, UK.

As expected, the serovar profile also varies among MSs within each source. In broiler carcasses, *S.* Enteritidis was isolated in 15 out of 23 countries where positive broiler samples were detected. *S.* Infantis was observed in 15 countries, and *S.* Typhimurium in 10. In the Czech Republic, Lithuania and Sweden, *S.* Agona predominated, while S. Kentucky was the most frequent serovar in Ireland, Malta and the United Kingdom. In Hungary, over 80% of all isolates were *S.* Infantis (Figure 5).

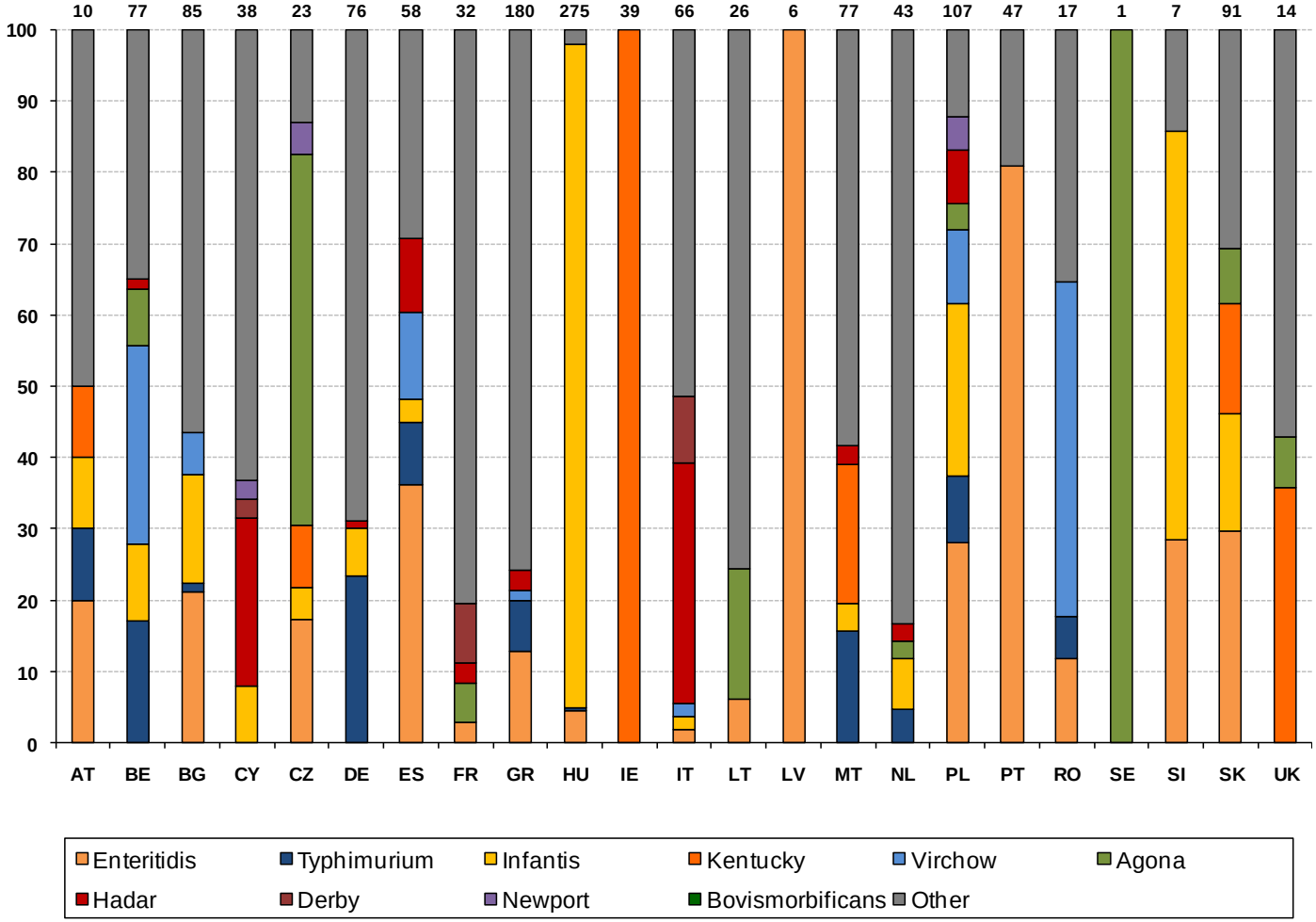


Figure 5. Relative frequency of selected *Salmonella* serovars in broiler carcasses. The number of positive samples is shown at the top of the bars.

In pigs, *S.* Typhimurium was observed in all countries with positive samples, followed by *S.* Derby and *S.* Enteritidis, which occurred in 19 and 20 out of the 23 countries, respectively. Those were also the serovars observed in larger proportions in the countries where they occurred (Figure 6).

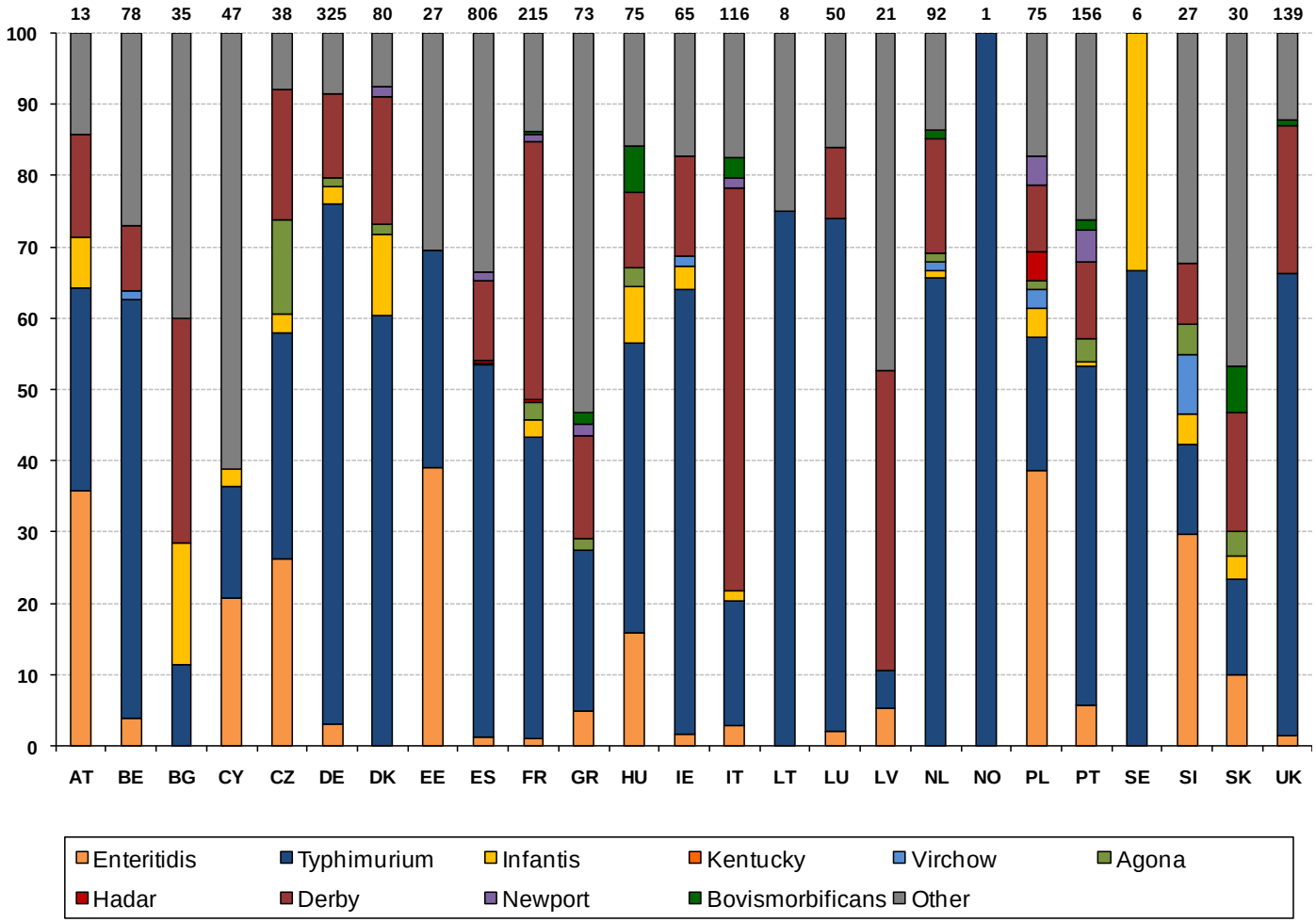


Figure 6. Relative frequency of selected *Salmonella* serovars in pigs. The number of positive samples is shown at the top of the bars.

In turkeys, *S*. Typhimurium, *S*. Derby and *S*. Hadar prevailed among the 11 selected serovars, with the exception of Slovenia and Hungary, where *S*. Infantis and *S*. Enteritidis were more frequent. However, most positives were among serovars aggregated as "Others", due to the importance of *S*. Saintpaul and *S*. Kottbus in this reservoir (Figure 7).

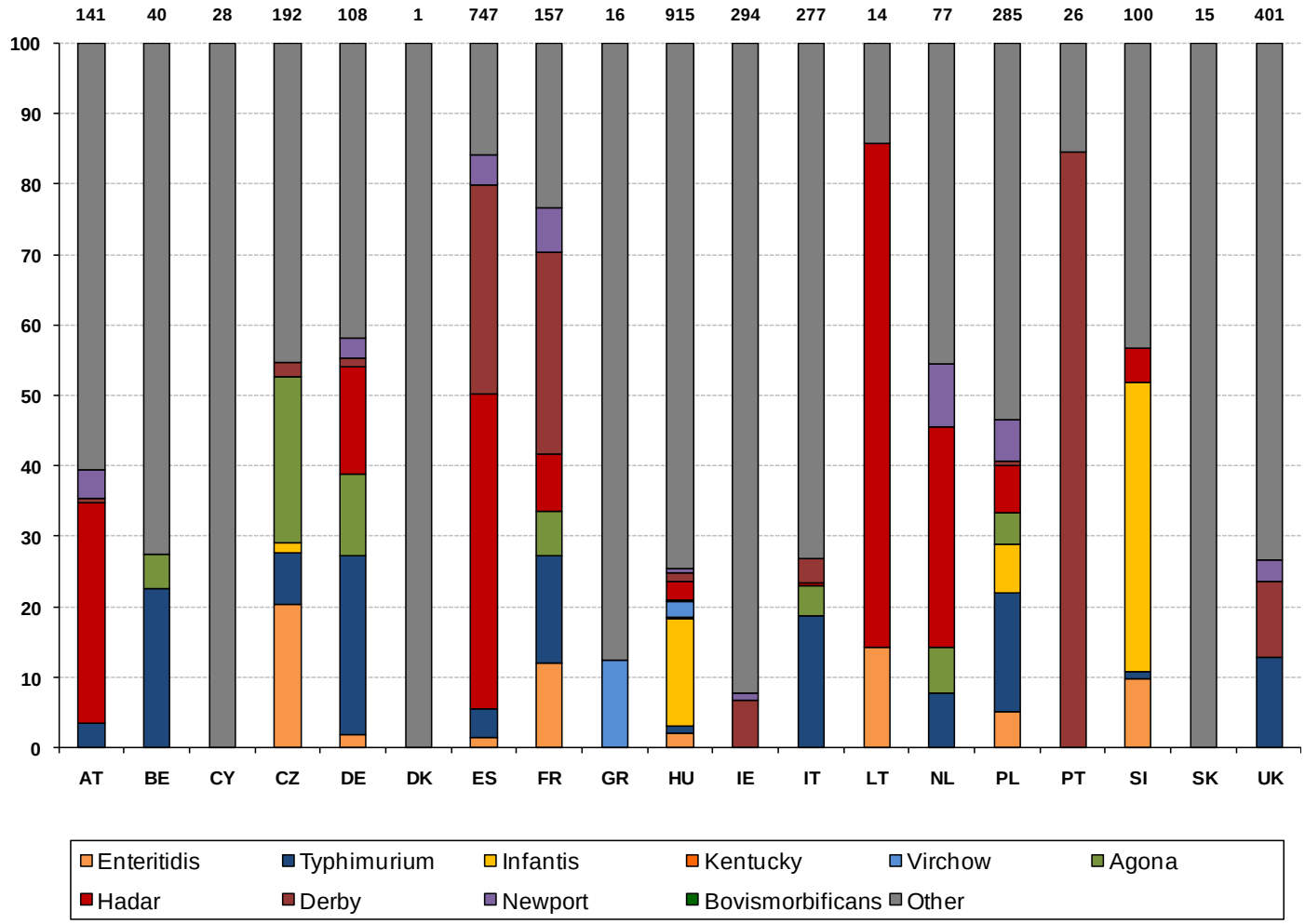


Figure 7. Relative frequency of selected *Salmonella* serovars in turkeys. The number of positive samples is shown at the top of the bars.

In layers, *S*. Enteritidis was present in 17 out of 22 countries, being the most frequent serovar in the majority of countries where it was detected. Finland, Luxembourg and Sweden were exceptions, with a predominance of *S*. Typhimurium. However, this could be due to a very small number of positive samples (one, one and five, respectively), the same occurring with *S*. Derby in Ireland (one out of two positives) (Figure 8).

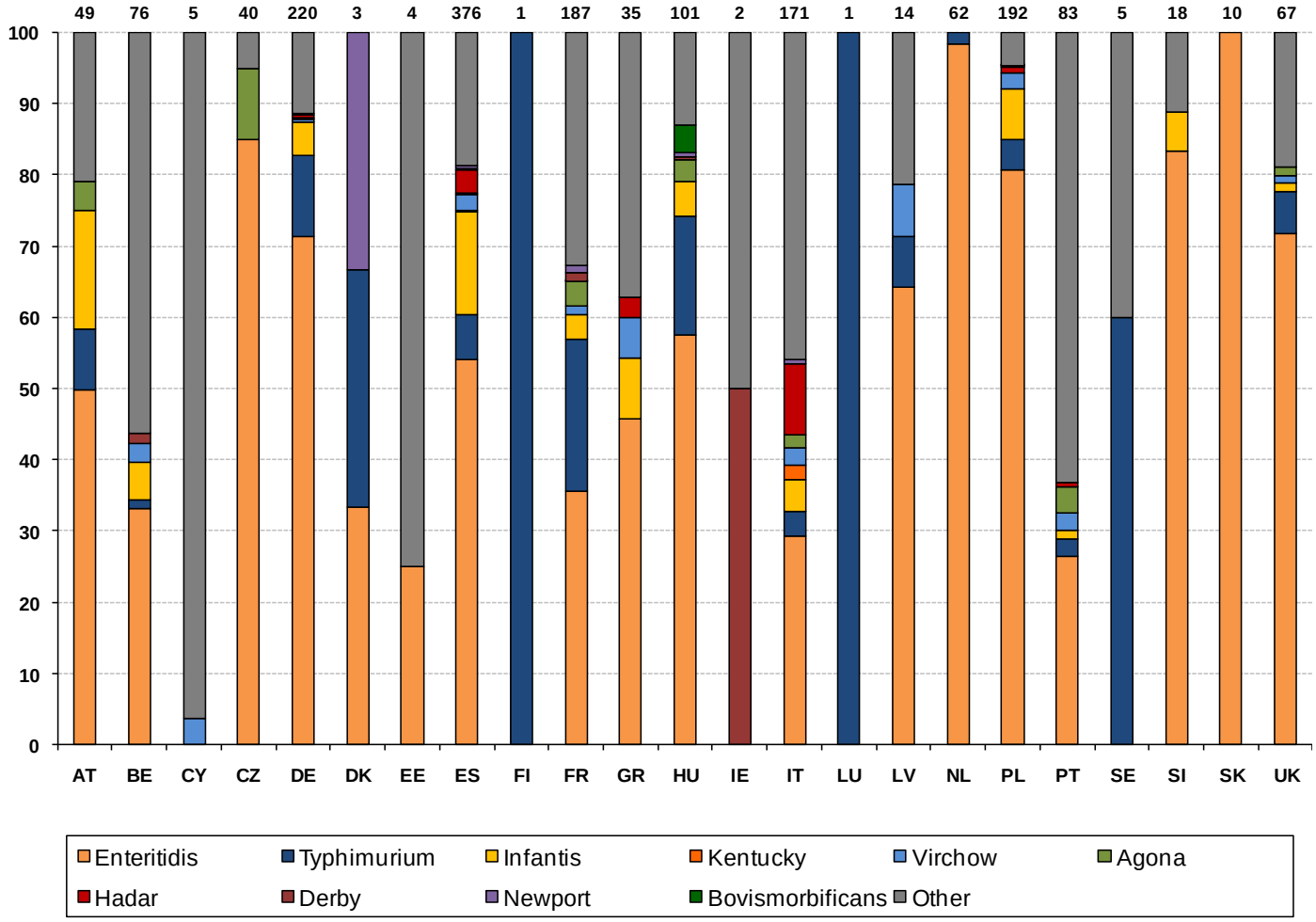


Figure 8. Relative frequency of selected *Salmonella* serovars in laying hens. The number of positive samples is shown at the top of the bars.

In cattle, *S*. Dublin (grouped as "Others" in Figure 9) was the dominant serovar in eight out of 22 countries, followed by *S*. Typhimurium, which was observed in large proportions in 12 countries and predominated in six. Other serovars of importance in specific countries in this source were *S*. Montevideo, *S*. Mbandaka and *S*. Infantis.

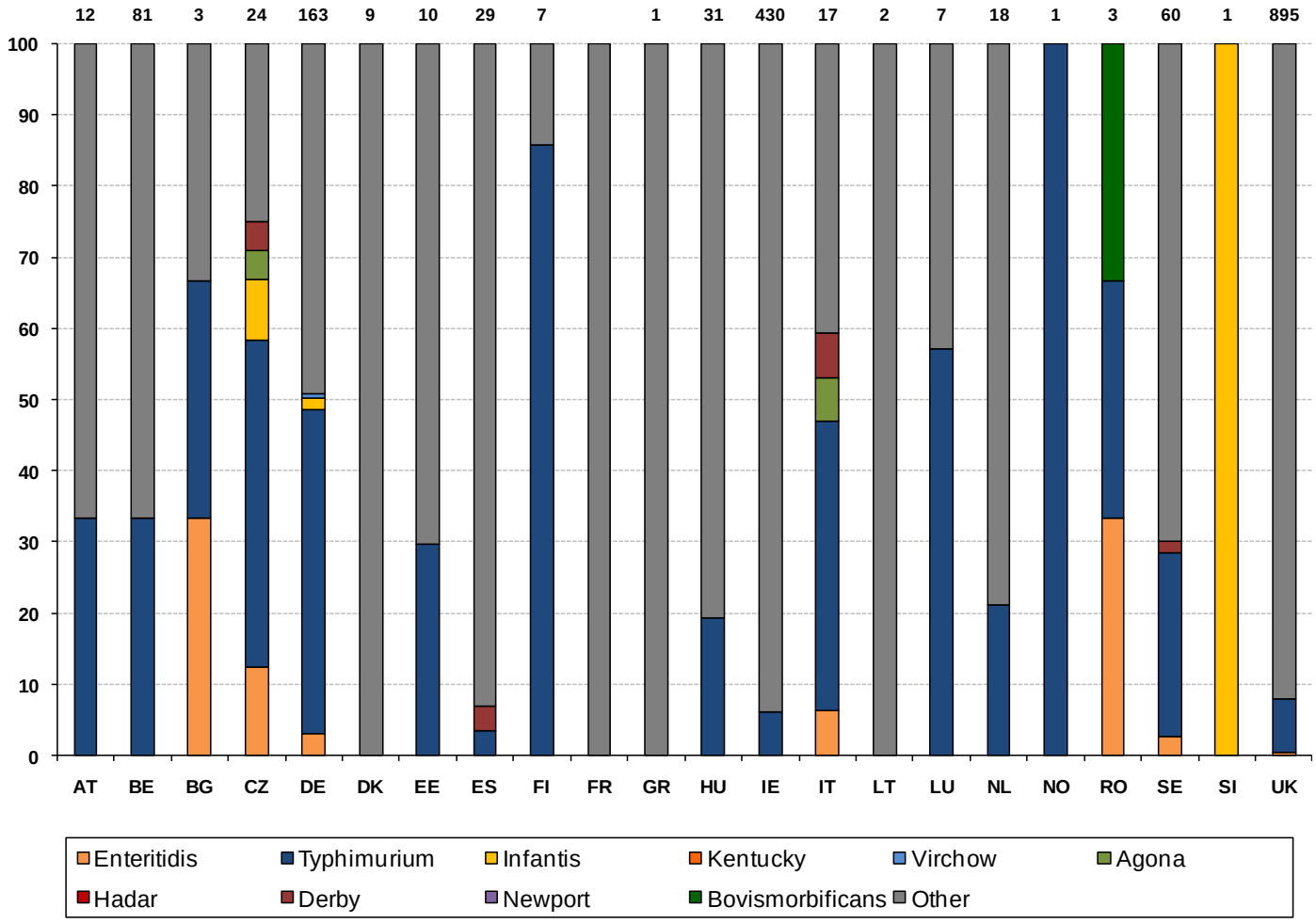


Figure 9. Relative frequency of selected *Salmonella* serovars in cattle. The number of positive samples is shown at the top of the bars.

Overall, *S.* Enteritidis and *S.* Typhimurium were the most frequent and widespread serovars in animal reservoirs, followed by *S.* Infantis, *S.* Hadar and *S.* Derby. An exception can be noted for turkeys, where *S.* Bredeney, *S.* Hadar and *S.* Saintpaul were more frequently isolated. Despite its importance in humans, *S.* Stanley was absent in all five animal sources.

Phage typing information was scarce and non-representative in the animal sources, as it is not mandatory for countries to report them. This kind of information was provided by seven out of 26 countries in the pigs BS; six out of 26 in the turkeys BS; only Italy and the Netherlands for laying hens, and no phage types were reported for broilers.

5.1.2.3. Food production and trade data

For consumption data calculations, it was assumed that the amount of food available in a country can be derived from how much is produced, how much is imported and how much is exported, thus making it necessary to obtain country-specific production data, as well as country-to-country imports and exports. By doing this, the model can take into consideration the amount of food present in a given country which originated from other countries and use the country- and food-specific serovar prevalences for the attribution. For this study, extra-EU food trade was not taken into consideration.

The statistical office of the European Union (EUROSTAT) was established in 1953, and its task is to provide the European Union with statistics at European level that enable comparisons between countries and regions. Among other information, it collects data on the production and trade of food products and animals for slaughter. International trade statistics, as produced by EUROSTAT, report the value and quantity of goods traded between EU MSs (Intrastat) and by EU MSs with third countries (Extrastat). European Community legislation ensures that the statistics provided to EUROSTAT by the MSs are based on legal texts and on harmonized definitions and procedures. However, an evaluation of the quality of the trade data collected by EUROSTAT has revealed major and persistent inconsistencies in the various MSs intra-EU trade statistics (EFSA, 2010b). Data availability varies depending on country and products selected, since the information is provided directly by MS, being subject to variations in national focus and cultural differences.

Food production data were derived by EFSA from the EUROSTAT database on slaughtered animals for food consumption and the EUROSTAT PRODCOM database.

5.1.2.3.1. Data selection and handling

The domestic amount of a product available for consumption in a country was estimated as Domestic Production minus Export, whereas the amount of imported food available for consumption in MS A originating from MS B was estimated as Import minus Re-export (when re-export was relevant). Due to differences between numbers reported in the production, imports and exports datasets, this operation in some cases resulted in negative amounts of national production available, meaning the volume exported was larger than what was domestically produced. In order to ensure that MSs would still have nationally produced food available in their own country, it was assumed that imported products could also be re-exported.The resulting trade matrix for each food source indicating the quantity transferred from an exporting to an importing country was used as input to the model.

This approach assumes that: 1) all the food available for consumption in the country is consumed; 2) countries do not export the whole national production of a food item 3) food exported by a country is not re-imported.

Dealing with missing data

Information on poultry for meat production (no differentiation between broilers and turkeys) were not available for Belgium in 2007 and 2008. Availability of data on the annual quantities of poultry, pork and bovine meat and eggs produced varied per year and per MS. For example, egg production data were lacking for several countries, and data for most food sources and most years were missing in some countries (e.g. Cyprus). Data on the export of eggs were not available for Cyprus. All MSs reported imports from other MSs for all food products in the study period.

Missing data on annual quantities of poultry meat products sold per MS, with differentiation between boilers, turkeys and other poultry species were obtained from the 2009 annual report of the Association of Poultry Processors and Poultry Trade in the EU Countries (AVEC, 2009).

Countries which had information missing for a year had the missing value estimated based on the percentage of increase or decrease between available years; when data from only one year was available, that value was used as surrogate for the missing years.

Dataset validation

The data obtained were validated by comparing it with consumption data available from the World Health Organization Global Environment Monitoring System Food Consumption Cluster Diets (GEMS/Food, 2006). The WHO data is available in grams/person/day, so the EUROSTAT data were converted to grams and divided by the country population (WHO, 2011) and by 365 to match the same unit. As the WHO data only offered the broad category "poultry", broilers and turkeys derived from the AVEC data were added together for this exercise. Relative proportions of consumption of poultry, pork and eggs were calculated, and a Proportional Similarity Index (PSI /Czekanowsky index) was calculated to compare those proportions between the two groups in each country. The PSI is an estimate of the area of intersection between two frequency distributions (Rosef, 2005), and is calculated as

$$PSI= 1-0.5*\sum|p_1-q_1| = \sum \min(p_1,q_1)$$

The method is traditionally used for calculating niche overlap and resource availability in population ecology (Feinsinger, 1981) or proportions of identified bacterial strains in epidemiology (Müllner, 2009; Müllner, 2010), but here we considered that each of the relative proportions among the three sources corresponds to the area under a probability curve, and so the same measure could be applied. A PSI of 1 means a complete overlap, or 100% similarity. An "overall PSI" for the whole dataset was calculated by using 24 instead of 1 for the subtraction, which arithmetically corresponds to the average of the country PSI values. In Table 10 are the PSI values comparing the relative proportions of consumption of the selected sources according to GEMS/Food and EUROSTAT-based surrogates. Nineteen out of 24 countries had a PSI of 0.9 or higher and three were larger than 0.8, suggesting that the consumption profiles composed using EUROSTAT data are highly similar to the original GEMS/Food profiles for most countries. The exception is noted for Cyprus, which may have an impact on the attribution estimates for this country. This issue is further addressed in the Discussion.

Table 10. Comparison of the relative proportion of consumption of pork, poultry meat and table eggs in the WHO GEMS/Food data and the surrogate values calculated from EUROSTAT data.

Country	WHO GEMS/Food (%)			EUROSTAT (%)			PSI
	Poultry	Pig	Egg	Poultry	Pig	Egg	
AT	16.7	70,9	12,4	18,8	68,8	12,4	**0,98**
BE	32.3	50,5	17,2	28,7	58,1	13,2	**0,92**
CY	38.7	48,3	13,0	96,8	2,9	0,3	**0,42**
CZ	28.6	52,7	18,6	28,4	52,9	18,7	**1,00**
DE	17.4	67,0	15,6	24,1	63,2	12,7	**0,93**
DK	19.4	64,2	16,5	13,1	81,3	5,6	**0,83**
EE	33.5	47,6	18,8	33,4	49,7	16,9	**0,98**
ES	25.8	61,0	13,2	30,9	56,2	12,9	**0,95**
FI	25.8	58,7	15,5	24,5	49,9	25,6	**0,90**
FR	32.9	47,7	19,4	42,1	39,5	18,4	**0,91**
GR	31.5	53,1	15,4	33,2	47,9	18,9	**0,95**
HU	33.2	49,8	17,0	41,0	42,0	17,1	**0,92**
IE	36.3	54,7	9,0	40,9	45,7	13,4	**0,91**
IT	24.4	59,9	15,7	31,0	53,9	15,1	**0,93**
LT	24.6	51,4	23,9	30,7	51,1	18,2	**0,94**
LU	47.8	44,3	8,0	32,2	45,7	22,1	**0,84**
LV	30.3	44,7	25,0	33,6	43,0	23,4	**0,97**
NL	16.2	59,6	24,2	31,0	51,5	17,5	**0,85**
PL	23.8	61,7	14,5	31,3	56,6	12,0	**0,92**
PT	32.7	54,2	13,1	34,8	50,7	14,5	**0,97**
SE	20.9	61,3	17,8	22,3	58,6	19,1	**0,97**
SI	37.9	50,9	11,2	44,6	39,2	16,2	**0,88**
SK	36.5	45,8	17,7	28,2	48,7	23,1	**0,92**
UK	44.2	38,7	17,1	48,0	33,7	18,3	**0,95**
Overall PSI							**0.91**

5.2. Final dataset for the source attribution model

Based on data availability and quality, 24 countries were included in model: Austria, Belgium, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden, The Netherlands and the United Kingdom. Countries which were initially analyzed but were excluded from the final dataset were Bulgaria, which presented 100% of human cases without serovar detailing; Romania, which only participated in one BS and had not enough surrogate data to be retrieved from the EUSR, besides reporting 84% of cases without serovar information; Norway, which is not part of the EU and does not report to EUROSTAT and Switzerland, which is also not part of the EU, and does not report to EUROSTAT and, more importantly, to TESSy.

As some countries had only a few human cases reported each year, cases from 2007 to 2009 were added together to increase data robustness. Considering that year is not one of the model dimensions and

prevalences in reservoirs were obtained as cross sectional data from different time points, this has no negative impact on the results, as long as they are interpreted as representative of the period as a whole, not of each year individually.

The number of outbreak-related cases per serovar were subtracted from the total number of domestically acquired cases to estimate the number of sporadic cases, if this was not already done by the reporting country. Furthermore, one case was subtracted from each outbreak and added to the sporadic cases, as it is assumed that the index case of an outbreak was a sporadic case.

Based on the availability of EU-wide homogeneous data or with at least good-quality surrogates, food-animal sources included were broilers, pigs, turkeys and laying hens (as the animal reservoir for eggs), and due to better completeness and availability, the resulting trade data from 2009 was used as surrogate consumption data for those sources. Data from the cattle reservoir were, in many occasions, well-representative of clinical cases, outbreak investigations or localized surveys. However, the need for a country-wide representative coverage that would allow cross-country comparisons was rarely met, and efforts to improve the dataset by using herd information from 2007-2008 or slaughterhouse carcass samples did not prove sufficient to obtain a representative dataset for this source in the model.

Serovar was chosen as subtyping level, and twenty-two of them were selected to be specifically addressed, based on their presence and importance in humans and in the main animal reservoirs: *S.* Agona, *S.* Anatum, *S.* Bovismorbificans, *S.* Braenderup, *S.* Brandenburg, *S.* Bredeney, *S.* Derby, *S.* Enteritidis, *S.* Hadar, *S.*Heidelberg, *S.* Infantis, *S.* Kentucky, *S.* Kottbus, *S.* Livingstone, *S.* London, *S.* Mbandaka, *S.* Montevideo, *S.* Newport, *S.* Rissen, *S.* Saintpaul, *S.* Typhimurium and *S.* Virchow. Albeit important in humans in most of the 24 countries, *S.* Dublin, *S.* Ohio and *S.* Stanley were not included in the list because *S.* Stanley was not isolated from the animal sources considered for the source attribution model, and the other two became irrelevant after the cattle reservoir was removed. For modeling purposes, serovars not included in the above list were aggregated as "Others". The building structure of the final *Salmonella* dataset (trade data not included) is shown in Figure 10.

Data were stored and analyzed in SAS Enterprise Guide, SAS Institute, SAS/STAT® User's Guide, Version 8, Cary, NC: SAS Institute Inc., 1999.

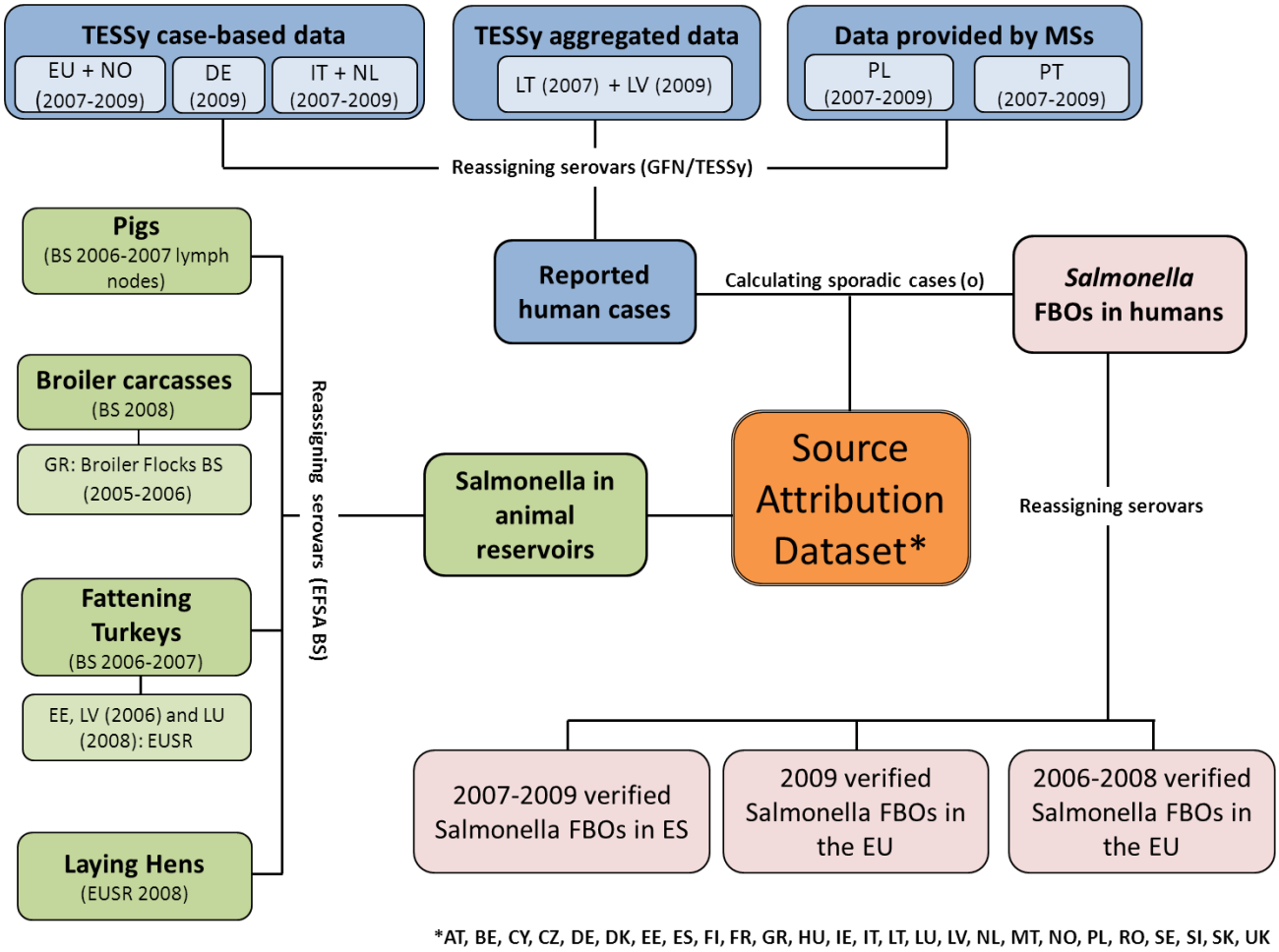


Figure 10. Diagram illustrating the construction of the final dataset for source attribution. Dark blue, dark green, pink and orange blocks represents datasets. Light blue blocks represent primary datasets originally provided to compose the blue blocks. Light green blocks represent surrogate data from other datasets used to complement the green blocks.

5.3. The European Union model (Manuscript II)

Before describing the methods used in this thesis, it is necessary to revise a few concepts.

5.3.1. Bayesian inference using Markov Chain Monte Carlo simulations

The probability of occurrence of an event can be described by a probability model. Those models typically contain parameters, which are regarded as fixed quantities that need to be estimated (King et al., 2010). The classical approach to estimate those values is to form a likelihood, which is a function of the model parameters, and to use the maximum likelihood estimate as the obtained value. Classical inference sees probability models as the data observed given the parameter, while Bayesian inference also considers the parameter given the data. This means that, in the Bayesian approach, model parameters are treated in the same way as the data, thus regarded as having distributions. Before data are entered into the model, those distributions are used as information about the parameters and described as *prior distributions*. So, a single parameter value is considered as only one of the possible values of that parameter, with its probability being defined by the prior distribution. The *posterior distributions* of the parameters can then be estimated, taking into account the prior information and the data (Lawson, 2009; King et al., 2010).

The choice of a prior distribution will depend on the prior knowledge available. There are three ways to choose a prior: a) subjectively, when there is no historical data available. In this case, the distribution shape and parameters express the experimenter's own personal experience and assumptions; b) objectively, when historical data on the distribution of parameter values or data from experiments done prior to the one being undertaken are available; c) when no strong preferences over values exist for some parameters, it is possible to assume *flat* (also known as *non-informative*) prior distributions. Uniform or close-to-uniform distributions are used, as they have a relatively flat shape, and so have little impact on the posterior distributions when compared to the likelihood of the data (Lawson, 2009).

The choice of a non-informative prior distribution can be made with some general understanding of the range and behavior of the variable. This makes it possible to estimate values for which no prior information is available, but which should be calculated relatively to each other. As an example, taking the following situation:

- 50% of 30 rat feed pellets are inoculated with 10 cfus of *S*. Enteritidis. This set of pellets will be called A;
- 50% of 30 rat feed pellets identical to set A are inoculated with 10 cfus of *S*. Kentucky. This set of pellets will be called B;
- Set A is fed to a group of 100 Wistar rats. This group will be called X;
- Set B is fed to a group of 100 Wistar rats identical to group X. This group will be called Y;
- It is assumed that each rat in groups X and Y eats the same amount of feed as the others, and that all 30 pellets from each set were consumed;
- It is assumed that all conditions for both groups, except the noted differences in exposure, were identical;
- 20 rats in group X become ill;
- 80 rats in group Y become ill;

The situation shows that some virulence factor (or factors) related to the two subtypes of *Salmonella* makes *S*. Kentucky more capable of causing disease in a population than *S*. Enteritidis. However, although this factor may be known on molecular level, it is not currently *quantifiable*. The scenario in group A, in

which 30 pellets with 50% contamination by *S.* Enteritidis generated 20 ill rats and the similar scenario in group B can be written as:

A) 0.5 * 30 = 20 cases
B) 0.5 * 30 = 80 cases

Considering that the two operations look identical but produce different number of cases, the virulence factor "q" should be entered, in order for them to make sense:

A) 0.5 * 30 * $q_{Enteritidis}$ = 20 cases
B) 0.5 * 30 * $q_{Kentucky}$ = 80 cases

Even if no real (measurable) values are available for $q_{Enteritidis}$ and $q_{Kentucky,}$ or if they in truth represent a group of factors instead of one, it is still possible to estimate them relatively to each other, by assuming a value of 1 for $q_{Enteritidis}$ and defining "q" as uniform distribution ranging from zero to 100. The value for $q_{Kentucky}$ will then be estimated relatively to 1 with a maximum possible value of 100, and it is expected to be larger than $q_{Enteritidis}$, as it causes more cases under the same conditions.

As explicit analytic forms for posterior distributions are usually not available, the approach used is to employ simulation procedures which result in samples from those distributions. The process of summarizing samples from a probability density as a way to integrate it is known as Monte Carlo Integration. It is based on the assumption that, if the density generates the sample, from that sample it is possible to approximately recreate the density (Smith and Gelfand, 1992). As an example, Figure 10 shows a distribution from which samples (blue dots) are taken. Those samples are then used to build a histogram, and the original curve can be derived again from the histogram shape. The values used as "results" of the model are summary statistics of the resulting sample, represented in the figure by the histogram. The most common is to use the mode as the most likely value, and dispersion measures as a way to assess the uncertainty around the values obtained (King et al., 2010) (Figure 11).

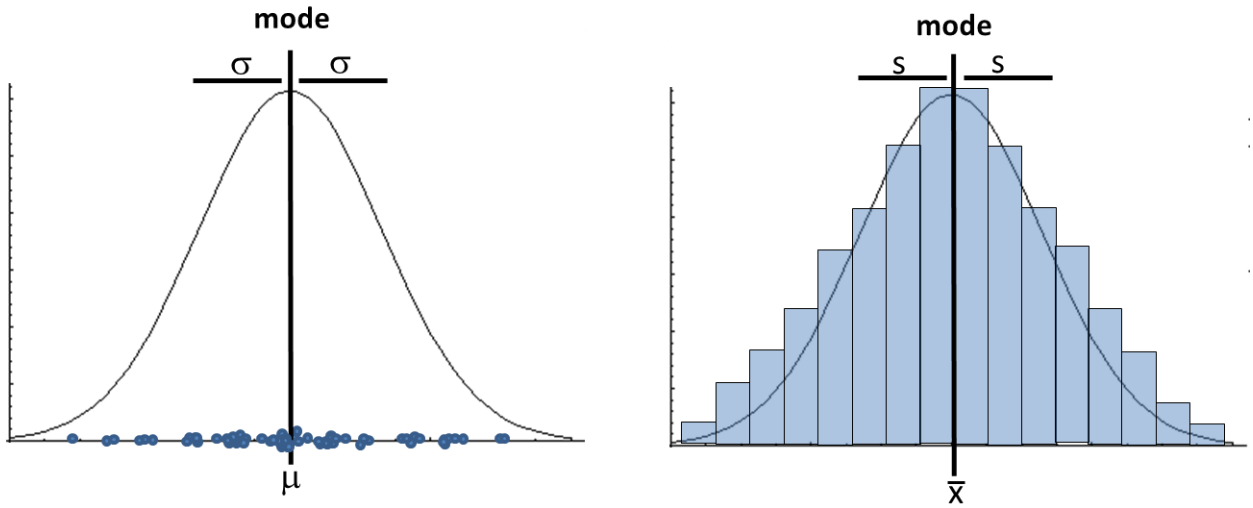


Figure 11. Monte Carlo Integration. If a density generates a sample, it is possible to approximately recreate that density from that same sample.

The most common way of generating samples from the posterior distribution for a Monte Carlo Integration is by using Markov Chains. Markov Chains are stochastic sequences of numbers in which each value in the sequence depends only on the last (Gilks et al., 1996). The nature of the Markov Chain states that it will tend to converge to a stationary distribution after a period, so, by assuming that the stationary distribution of the chain is equal to the posterior distribution of the model, the set of values generated will tend to converge to the general shape and specifications of the posterior distribution (King et al., 2010). The whole process is known as Markov Chain Monte Carlo (MCMC), and the generation of each sample is called an *iteration*.

In order to construct a MCMC sampler, the assumption that the stationary distribution of the chain is the posterior distribution of the model must be fulfilled. For that, the Gibbs sampler, which is a variation of the Metropolis-Hastings rejection algorithm, can be used to define the desired stationary distribution and reject generated values that do not fit it (Gilks et al., 1996; King et al., 2010). Given the principle by which Markov Chains are generated, each value generated is closely correlated to the one generated immediately before. When the correlation gets too strong, the chain may "double back" and start generating numbers that have already been drawn, and so the sampling stops progressing (Figure 12).

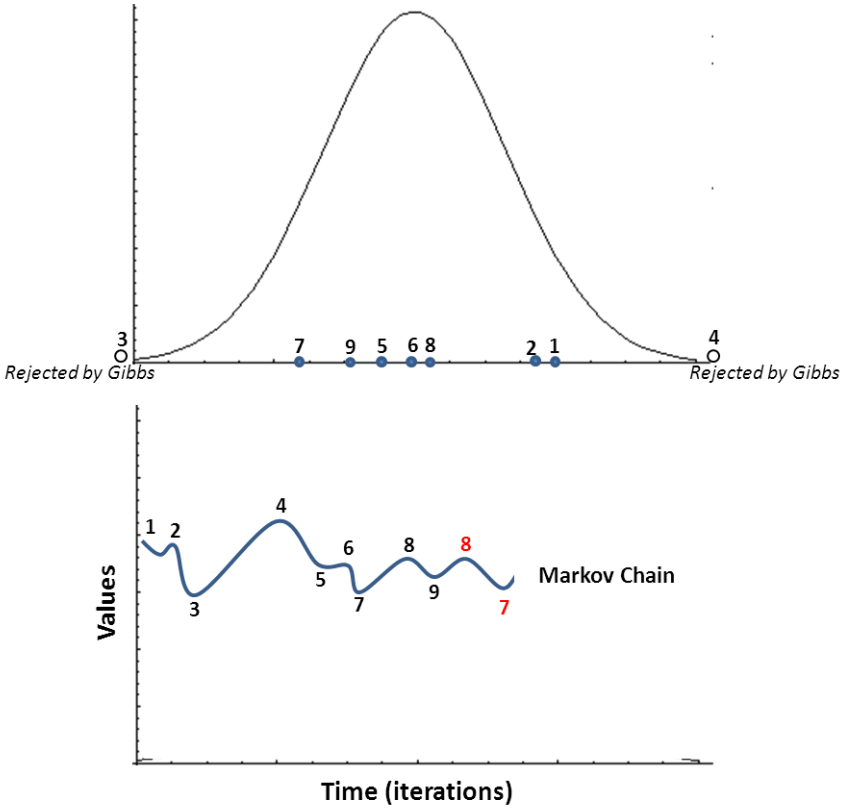


Figure 12. Markov Chain doubling back while sampling from a posterior distribution.

To avoid that, it is possible to define that a certain amount of values must be generated and "pre-selected", but only giving final acceptance to the value with the smallest correlation coefficient in relation to the others. This method is known as over-relaxation, and it also helps maintaining the randomness of the sample, as it allows the sampling to move from the area of the posterior distribution that was being sampled to another area.

When the chain reaches the target distribution, it is said that the model *converged*. Considering that the stationary distribution has been predefined, and that samplers which do not allow the values drawn to deviate from the chosen specifications have been used, it is necessary to re-assert the randomness of the process. This is partially achieved through over-relaxation, but mostly by the simultaneous running of more than one Markov Chain, with starting values located at different points of the distribution. As all chains have the same posterior distribution defined as their stationary distribution, it is expected that all of them converge to the same value range at some point.

Model convergence can be monitored as described by Gelman and Rubin (1992), by observing that the variance between chains should not be larger than within-chain variance. The two variances are compared to generate the Potential Scale Reduction Factor (PSRF). A PSRF close to one indicates that approximate convergence has been reached (van Valkenhoef et al., 2012). This can be visually assessed in a Brooks-Gelman-Rubin (BGR) diagram, where the PSRF is plotted against the number of iterations run (Figure 13).

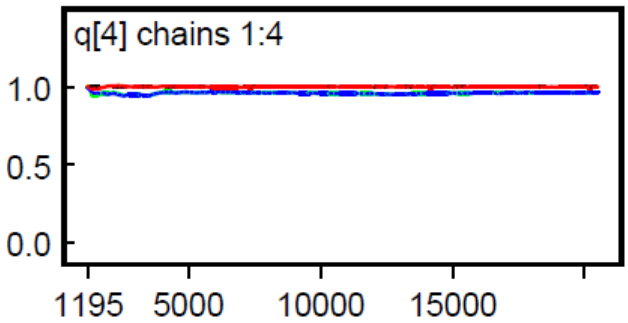


Figure 13. BGR diagram from WinBUGS showing four chains converging between iterations 2000 and 5000.

Convergence can also be checked visually in a time-series graph, where values generated are plotted against the number of iterations run; in this graph, lines representing the chains should overlap and be reasonably stable to consider that convergence has occurred (Figure 14).

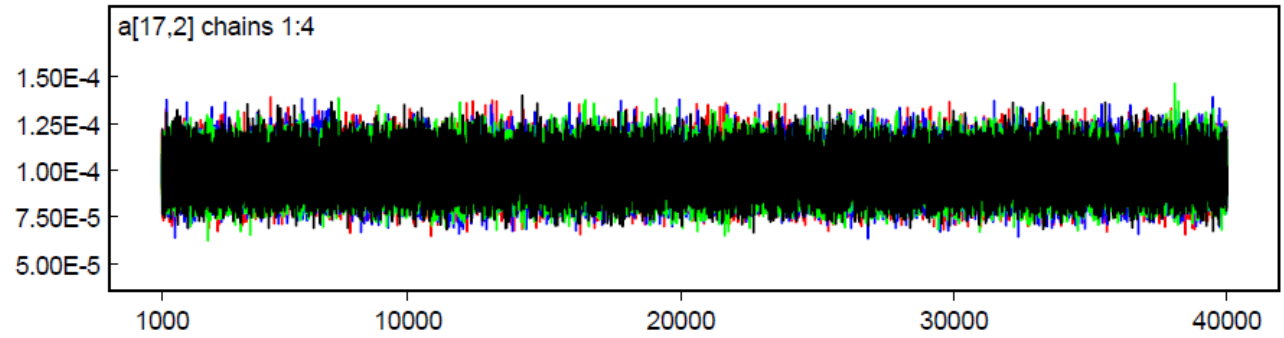


Figure 14. Time-series diagram from WinBUGS showing overlapping of four chains (shown in red, blue, green and black).

In both graphs, the evaluation of how much chains overlap or how stable they are is rather subjective, and so it is not possible to affirm that a model has certainly converged. However, situations in which convergence has certainly not occurred are more clearly identified, as individual chains are seen in different areas of the graph, or moving briskly between a wide range of values.

5.3.2. The EU model

The EU model was based on Hald et al. (2004) and Pires et al. (2010), and modified to 1) accommodate data from multiple countries, 2) use trade data as a surrogate for consumption data and 3) allow for the detection of the country of origin of the sources causing cases in the reporting countries.

The model is built on a Bayesian framework and includes three dimensions: the *Salmonella* subtype (*i*), the food/animal source (*j*) and the country of attribution (*c*). It attributes sporadic cases of salmonellosis to animal reservoirs, to international travel, and gathers cases related to sources and/or serovars not included in the model as "unknown".

A sporadic case is defined as a subject that could not be associated with a recognized foodborne disease outbreak. Outbreak-related cases are added to the final results of the model, being attributed to the source implicated in the outbreak, if that is known. If not, they are considered outbreaks with unknown source. The outbreak cases are not modeled together with the other cases, as outbreaks caused by serovars only occurring in one source will result in an overestimation of the total cases attributable to that source. At the same time, outbreaks of more ubiquitous serovars would underestimate the number of infections caused by that source. Cases which do not specifically report a history of travel up to one week prior to symptom onset are assumed to be domestic.

Underreporting of cases is taken into consideration by multiplying sporadic reported cases by the correspondent UF (Havelaar et al., 2012) in each MS after attribution. Outbreak cases are assumed to have been reported in a more complete manner, and so were not adjusted. The underreporting factors were fitted as lognormal distributions, following the methodology described in Hald et al. (2012). Values used to include UFs in the model are presented in Figure 15, extracted from Hald et al. (2012).

	Mean(Ln)	Sdev(Ln)
Austria	2.1	0.8
Belgium	0.9	0.9
Bulgaria	6.3	0.8
Cyprus	4.9	0.8
Czech Republic	3.1	0.8
Denmark	1.2	0.8
Estonia	2.5	0.8
Finland	-1.3	0.8
France	3	0.8
Germany	2	0.8
Greece	6.8	0.8
Hungary	3.9	0.8
Ireland	1.1	1.1
Italy	4	0.8
Latvia	3.5	0.8
Lithuania	3.8	0.8
Luxembourg	1	1
Malta	5.1	0.8
Poland	4.5	0.8
Portugal	7.4	0.8
Romania	5.5	0.8
Slovakia	3.7	0.8
Slovenia	3.4	0.9
Spain	5.1	0.8
Sweden	-1	0.8
The Netherlands	3	0.8
United Kingdom	1.7	0.8

Figure 15. Mean and standard deviation values of the fitted lognormal distributions for the UFs. Source: Hald et al. (2012).

5.3.2.1. Model parameters and specifications

The model takes into account the number of cases caused by a serovar, the prevalence of each serovar in each source in each country and the relative impact of a set of unknown factors, as described in Hald et al. (2004). Those factors were included as multi-parameter flat priors, and account for the differences in the ability of different subtypes to cause disease and of different sources to act as vehicles for infection. Multiple loops were included to accommodate data from the 24 countries. An overview of the model parameterization can be drawn as:

a_{cj} ~ Uniform (0,100)

q_i ~ Uniform (0,100)

λ_{ci} ~ Poisson (o_{ci}),

$\lambda_{ci} = \sum_{i=1}^{n} \sum_{k=1} \lambda_{ckji}$

$\lambda_{ckji} = p_{kij} * m_{ckj} * a_{cj} * q_i$

where: 1) λ_{ckji} is the expected number of cases per serovar *i* and source *j* reported in country *c* and caused by food produced in country *k*; 2) p_{kij} is the prevalence of serovar *i* in source *j* in country *k*; 3) m_{ckj} is the amount of source *j* available for consumption in country *c* produced in country *k*; when a source is domestically produced in the country of attribution, *c*=*k*; 4) a_{cj} is the source-dependent factor for source *j* in country *c*; 5) q_i is the subtype-dependent factor for serovar *i*. The source-dependent factor a_{cj} was assumed to vary between countries, accounting for variability in consumption patterns and preferences not captured by m_{ckj}, also including general variations between sources, *e.g*, bacterial load/concentration in the food and processing, handling or preparation practices.The subtype-dependent factor q_i is a one-dimensional parameter, meaning that it is a property of the *Salmonella* serovar and assumed independent of the country of infection.

The amount of food source available for consumption in the country where a *Salmonella* case was reported considers both domestically produced and imported foods (m_{ckj}). The q_i prior for *S.* Enteritidis is defined as 1, and all other q_i values are estimated relatively to this one. The number of human sporadic and domestic cases attributed to each source per country (λ_{cji}) is estimated assuming a Poisson distribution of the observed number of sporadic cases per subtype per country (o_{ci}). Model parameters are presented in Table 11.

The model was built in WinBUGS 1.4 (http://www.mrc-bsu.cam.ac.uk/bugs/), which uses Markov Chain Monte Carlo (MCMC) with Gibbs sampling as a default to obtain summary values for posterior distributions. Five independent chains ran for 40,000 iterations each to obtain the values for a_{cj} and q_i. Each chain had a different set of starting values for the priors, widely dispersed in the target distribution. Chain convergence was monitored using the methods described by Gelman and Rubin (Gelman and Rubin, 1992) and was considered to have occurred when the variance between the different chains was no larger than the variance within each individual chain, and when the chains had reached a stable level.

Table 11. Parameters used to estimate the number of sporadic cases of salmonellosis attributable to the animal sources

Notation	Description	Estimation
i (1-22)	*Salmonella* serovar	-
j (1-4)	Food-animal source	
c (1-24)	Country where the human case was reported	
k (1-24)	Country of origin of the food product[(a)]	
o_{ci}	Observed cases caused by serovar *i* in country c	Data
ob_{ci}	Observed cases caused by serovar *i* known to be outbreak related in country c. For each outbreak, one case was subtracted so that one outbreak contributed with one sporadic case.	Data
yt_{ci}	Observed cases caused by subtype *i* in country c that was reported as travel-related	Data
p_{kji}	Prevalence of subtype *i* in source *j* in country *k*	Data
m_{ckj}	Amount of source *j* available for consumption in country *c* produced in country *k*[(a)]	Data
a_{cj}	Source-dependent factor for source *j* and country c	dunif(0,maxa_{cj})
q_i	Subtype-dependent factor for subtype *i*	dunif(0,maxq_i)
uf_c	Underreporting factor for country c	dllnorm(μ,σ)
$spdo_{ci}$	Total number of sporadic cases caused by subtype *i* in country *c*	o_{ci} -yt_{ci} – (ob_{ci}+ 1)

(a) If the food is produced and consumed in the same country, c=k

6. RESULTS

6.1 Attribution of cases to food sources in the country of reporting (Manuscript II)

The total number of cases attributed after applying the UFs and a summary of the attributed parcels in each country are shown in Table 12.

Table 12. Proportion of *Salmonella* cases attributed to food sources, outbreaks and international travel in EU MSs and regions[(a)], 2007-2009.

Country	Attributed parcel (%)							Total attributed
	Broilers	Pigs	Turkeys	Layers	Outbreaks[(b)]	Travel	Unknown	
CZ	0.1	10.9	1.8	84.6	0.0	1.7	0.8	1,178,000
HU	4.5	26.7	5.4	54.9	0.2	0.2	8.1	1,172,000
PL	25.1	47.8	1.2	23.0	0.1	0.1	2.7	3,178,000
SK	0.0	18.0	2.6	76.8	0.0	0.8	1.7	1,051,000
Eastern EU[(a)] total	12.9	32.7	2.3	48.3	0.1	0.5	3.2	6,579,000
DK	3.5	18.0	19.6	10.1	6.8	23.7	18.3	26,331
EE	4.6	27.5	2.1	55.0	0.3	7.9	2.6	19,970
FI	0.7	4.7	1.6	2.4	5.9	80.1	4.6	3,210
IE	1.5	27.2	8.8	14.6	0.9	31.7	15.3	6,660
LT	1.2	9.5	0.7	86.9	0.1	0.3	1.2	448,600
SE	0.5	4.8	1.7	2.5	4.4	75.9	10.2	5,851
UK	0.6	11.7	10.1	35.5	0.0	24.3	17.8	276,400
Northern EU[(a)] total	1.1	11.4	4.3	66.0	0.3	9.8	7.1	787,021
AT	0.1	14.4	3.7	59.8	0.3	12.2	9.4	91,130
BE	2.3	74.2	9.2	2.9	0.1	0.0	11.2	40,600
DE	0.5	33.1	1.3	52.0	0.2	5.3	7.6	1,271,000
FR	13.4	34.3	12.6	2.9	0.2	0.0	36.5	492,000
LU	4.4	8.5	6.9	49.8	0.0	9.6	20.7	2,154
NL	4.6	27.3	9.7	26.2	0.5	14.2	17.5	96,580
Western EU[(a)] total	3.9	33.1	4.8	38.0	0.2	4.7	15.4	1,993,464
CY	4.8	51.1	6.4	8.9	0.0	3.8	24.9	87,240
LV	0.9	13.7	0.3	82.5	0.4	1.5	0.7	98,880
ES	0.1	33.1	12.9	43.1	0.0	0.0	10.7	2,627,000
GR	1.2	9.5	0.4	78.3	0.0	2.3	8.3	2,390,000
PT	42.3	36.3	0.6	9.1	0.0	0.4	11.4	3,206,000
IT	2.3	73.2	5.3	2.2	0.0	1.3	15.8	766,000
SI	0.5	20.6	4.0	59.5	0.6	0.0	14.7	104,600
Southern EU[(a)] total	15.4	31.5	4.5	36.8	0.0	0.9	10.9	9,279,720
EU total	12.6	31.1	3.8	42.4	0.1	1.6	8.5	27,998,690

(a) EU regions as defined by the United Nations. Eastern Europe: Czech Republic, Hungary, Poland and Slovakia. Northern Europe: Denmark, Estonia, Finland, Ireland, Latvia, Lithuania, Sweden and the United Kingdom. Southern Europe: Cyprus, Greece, Italy, Portugal, Slovenia, Spain. Western Europe: Austria, Belgium, France, Germany, Luxembourg and the Netherlands.

(b) The proportion of outbreak cases were derived directly from the reported data (i.e. they were not estimated and consequently no Credibility Intervals were calculated); includes outbreaks with unknown source. Outbreak cases for which the source was identified were assigned to the correspondent animal sources.

Detailed accounts of the number of cases attributed to each category and 95% Credibility Intervals for countries and regions are available in Appendices A and B. The most important source of human salmonellosis at the EU level was estimated to be the laying hen reservoir (i.e. eggs), with 42.4% of cases (7,903,000 cases, 95% Credibility Interval (CI) 4,181,000 – 14,510,000), followed by 31.1% attributed to pigs (5,800,000 cases, 95% CI 2,973,000 – 11,100,000). Broilers and turkeys were estimated to be less important sources of *Salmonella*, contributing with 12.6% (2,350,000 cases, 95% CI 736,300 – 6,194,000) and 3.8% (702,400 cases, 95% CI 325,500 – 1,590,000), respectively. A total of 1.6% (292,400 cases, 95% CI 150,700 – 562,700) of all salmonellosis cases were reported as being travel-related, and 0.1% (13,848) of cases were reported as being part of outbreaks with unknown source.

Of all *S*. Enteritidis infections, 63% (7,504,000 cases, 95% CI 3,964,000-13,770,000) were attributed to laying hens, whereas 90.8% of *S*. Typhimurium originated from pigs (2,950,000 cases, 95% CI 1,510,000-5,663,000). Compared to infections attributed to layers and pigs, a large proportion of cases were caused by other serovars in other sources, such as 4.5% *S*. Infantis in broilers (106,600 cases, 95% CI 32,560-284,500) and 9.2% *S*. Newport (226,296 cases, 95% CI 84,379-567,930) or 4.5% *S*. Saintpaul (33,580 cases, 95% CI 18,052-62,443) in turkeys. In those sources, these serovars were not the most frequently associated with cases, but still constituted a significant burden. The proportions attributed to the main serovars in each animal reservoir can be observed in Appendix C.

At regional level, layers were the most important source in all regions, with between 36.7% and 66.0% of the *Salmonella* reported cases attributed to this source. Pigs were the second most important source, notably in Eastern (32.7%), Western (33.1%) and Southern EU (31.5%). In Southern and Eastern EU, broilers were also an important source, with respectively 12.9% and 15.4% of cases. A large proportion of the reported *Salmonella* infections in Northern European countries were acquired abroad, when compared to the other regions, where foreign travel seemed to be of less importance.

When looking at the attributed proportions within specific countries, the laying hen reservoir was estimated as the most important source of salmonellosis in 13 countries (Austria, Czech Republic, Estonia, Germany, Greece, Hungary, Latvia, Lithuania, Luxembourg, Slovenia, Slovakia, Spain and the United Kingdom), whereas pigs were the larger contributor for salmonellosis in eight (Belgium, Cyprus, Finland, France, Ireland, Italy, Poland and Sweden); the proportion of disease attributed to layers and pigs were similar in the Netherlands. In Denmark, the most important food-animal source was estimated to be turkeys, and broilers were the major source in Portugal. In Finland and Sweden, the majority of *Salmonella* infections were estimated to be travel-related. Travel was also an important source in Ireland, the UK and Denmark, although to a lower extent. Figure 16 presents the relative proportion of cases attributed to animal sources, travels and outbreaks with unknown source.

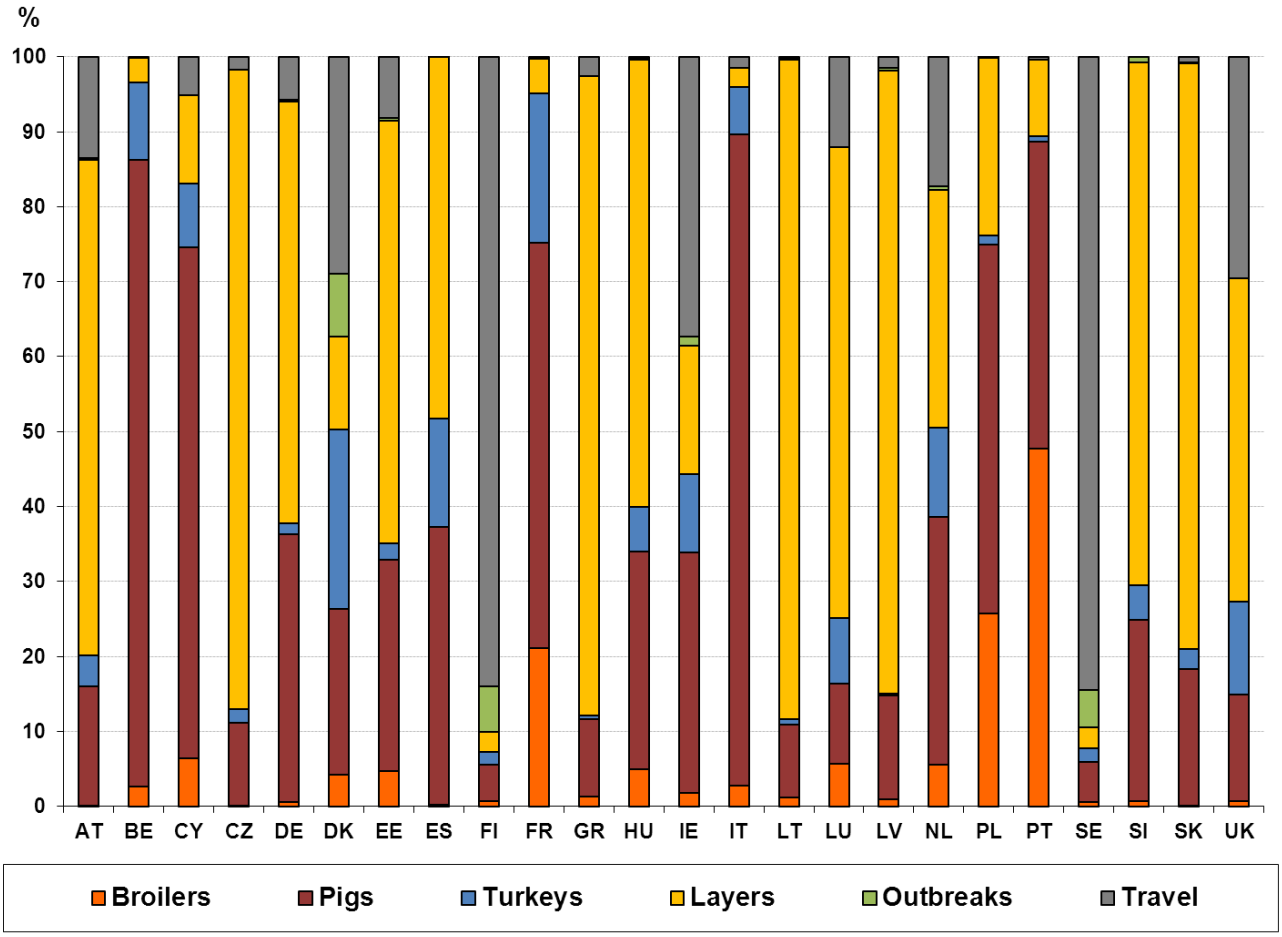


Figure 16. Proportion of *Salmonella* human cases attributed to food animal reservoirs, travel and outbreaks in 24 EU Member States, 2007-2009 (median %).

6.2. Attribution of cases in the EU to countries of origin of the food sources

As mentioned earlier, a new feature of this model is the ability to estimate the country of origin of cases attributed in other countries, as country-specific prevalences and amounts are used. For all graphs shown, cases reported in the country of origin are also included in the total; as an example, Portuguese cases attributed to broilers are among the total cases attributed to Portuguese broilers in Figure 18c.

When considering all sources together, Poland was estimated to be the most important source-country for human salmonellosis in the EU, contributing with 21.3% of cases (3,563,710 cases, 95% CI 911,750 – 10,818,900), followed by 18.4 from Spain (3,081,090 cases, 95% CI 898,170 – 9,056,800) and 14.5 from Portugal (2,422,142 cases, 95% CI 361,368 – 8,508,397) (Figure 17). Country-specific estimates with 95% Credibility Intervals are shown in Appendix D.

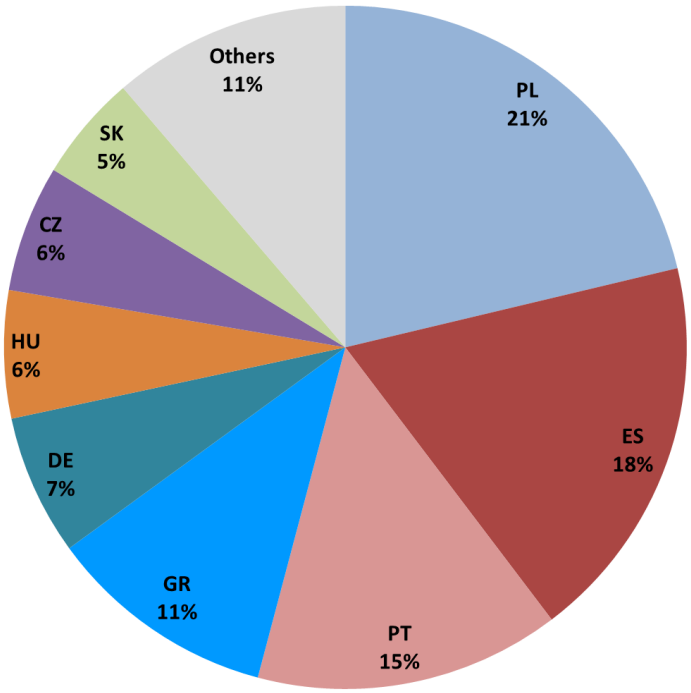


Figure 17: Proportion of cases of *Salmonellosis* in the EU originating from sources of each MSs.

When looking at specific reservoirs, among cases attributed to layers 21.5% (1,701,000 cases, 95% CI 256,400 – 5,944,000) were estimated to have originated from Greece; 17.9% (1,414,000 cases, 95% CI 406,000 – 4,286,000) from Spain, 16.3% (1,287,000 cases, 95% CI 492,000 – 3,162,000) from Poland and 11.1% (874,200 cases, 95% CI 142,000 – 1,299,000) from the Czech Republic (Figure 18a). Pork-attributed cases were estimated to originate mostly from Poland (24.2% or 1,402,000 cases, 95% CI 257,000 – 4,721,000), Spain (22.5% or 1,306,000 cases, 95% CI 423,700 – 3,556,000) and Portugal (15.1% or 876,000 cases, 95% CI 134,800 – 3,040,000) (Figure 18b). Portuguese broilers were responsible for 55.6% of cases (1,305,000 cases, 95% CI 198,500 – 4,535,000) attributed to that source in the EU. Poland was the second most important contributor, with 34.2% of cases (803,600 cases, 95% CI 131,400 – 2,768,000) (Figure 18c). Cases attributed to Turkeys originated mostly from Spain (43.1% or 302,600 cases, 95% CI 55,350 – 1,029,000), with large contributions from France (16.6% or 116,700 cases, 95% CI 43,460 – 287,300), Hungary (12.0% or 84,060 cases, 95% CI 27,580 – 230,500) and Poland (10.1% or 71,110 cases, 95% CI 30,950 – 167,900) (Figure 18d). Number and percentage of cases attributed to sources from individual countries are show in Appendices D and E, respectively.

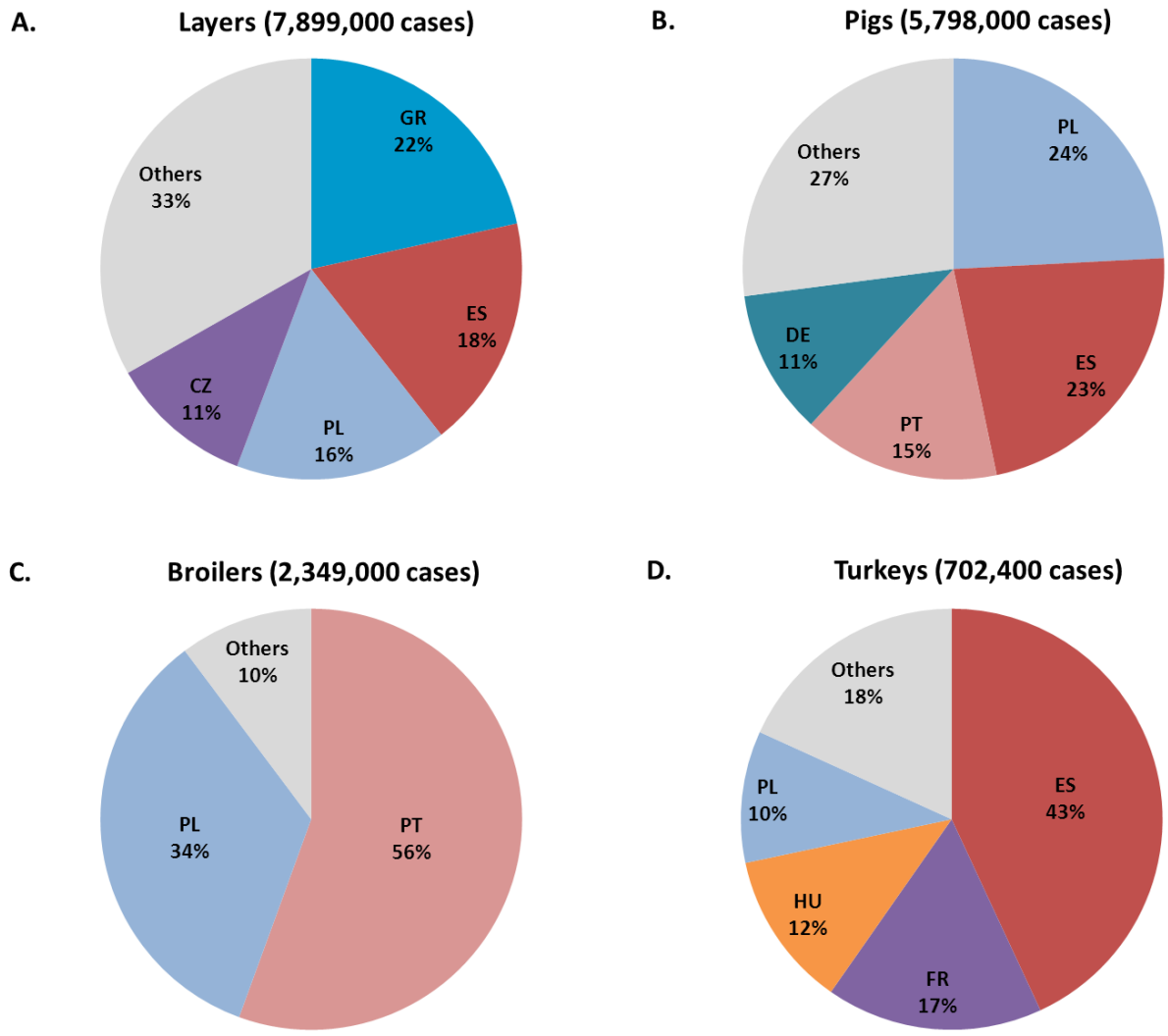


Figure 18: Proportion of country of origin of food-sources to which cases of *Salmonellosis* were attributed.

6.3. Posterior estimates for a_{cj} and q_i

The estimated ability of food sources to act as a vehicle for disease (a_{cj}) was higher for layers in 11 countries (Austria, Czech Republic, Estonia, Germany, Greece, Hungary, Lithuania, Luxembourg, Latvia, Slovenia and Slovakia) and turkeys in 10 countries (Belgium, Cyprus, Denmark, Finland, France, Ireland, the Netherlands, Spain, Sweden and the UK). In Italy and Poland, the highest a_{cj} was estimated for pigs, whereas in Portugal results revealed a higher estimate for broilers. Values estimated for a_{cj} are shown in Table 13.

Regarding the ability of different serovars to cause disease, the highest q_i value was estimated for *S.* Kentucky, followed by *S.* Enteritidis (value fixed to 1), *S.* Newport, *S.* Virchow and *S.* Typhimurium. Estimated values for q_i for all serovars are shown in Table 14.

Table 13. Estimated values for a_{cj}, source-dependent factor (mean and 95% Credibility Interval)

Country	Broilers	95% CI		Pigs	95% CI		Turkeys	95% CI		Layers	95% CI	
AT	5.38E-07	[2.00E-08 ,	2.85E-06]	4.93E-05	[4.30E-05 ,	5.58E-05]	1.39E-04	[1.07E-04 ,	1.76E-04]	7.42E-04	[7.14E-04 ,	7.70E-04]
BE	2.69E-05	[2.10E-05 ,	3.36E-05]	2.00E-04	[1.89E-04 ,	2.11E-04]	5.27E-04	[4.54E-04 ,	6.06E-04]	1.73E-05	[1.02E-05 ,	2.48E-05]
CY	1.66E-06	[8.27E-07 ,	2.74E-06]	3.51E-05	[2.93E-05 ,	4.12E-05]	1.39E-03	[4.90E-04 ,	2.59E-03]	1.03E-04	[4.97E-05 ,	1.73E-04]
CZ	1.81E-06	[3.07E-07 ,	3.86E-06]	1.41E-04	[1.30E-04 ,	1.52E-04]	6.25E-04	[4.98E-04 ,	7.67E-04]	8.63E-04	[8.52E-04 ,	8.74E-04]
DE	1.11E-05	[2.36E-06 ,	2.13E-05]	1.13E-04	[1.07E-04 ,	1.19E-04]	1.05E-04	[8.92E-05 ,	1.22E-04]	6.08E-04	[6.02E-04 ,	6.15E-04]
DK	9.31E-05	[5.06E-05 ,	1.42E-04]	2.19E-05	[1.91E-05 ,	2.48E-05]	2.39E-03	[2.01E-03 ,	2.79E-03]	1.08E-04	[8.54E-05 ,	1.31E-04]
EE	7.53E-05	[1.54E-05 ,	1.55E-04]	9.84E-05	[7.51E-05 ,	1.24E-04]	3.43E-04	[1.27E-04 ,	6.85E-04]	4.48E-04	[3.92E-04 ,	5.04E-04]
ES	4.12E-08	[1.54E-09 ,	2.18E-07]	1.07E-05	[9.91E-06 ,	1.15E-05]	1.54E-04	[1.33E-04 ,	1.76E-04]	1.01E-05	[9.78E-06 ,	1.04E-05]
FI	6.27E-04	[3.15E-05 ,	1.98E-03]	4.70E-04	[3.35E-04 ,	5.98E-04]	2.71E-03	[7.36E-04 ,	4.99E-03]	8.39E-06	[3.53E-06 ,	1.34E-05]
FR	5.10E-05	[4.73E-05 ,	5.50E-05]	3.86E-05	[3.61E-05 ,	4.10E-05]	1.47E-04	[1.33E-04 ,	1.62E-04]	5.99E-06	[4.65E-06 ,	7.50E-06]
GR	6.83E-07	[2.51E-08 ,	3.53E-06]	8.43E-06	[6.90E-06 ,	1.01E-05]	1.07E-05	[1.51E-06 ,	3.47E-05]	2.61E-05	[2.45E-05 ,	2.75E-05]
HU	4.93E-06	[4.23E-06 ,	5.67E-06]	1.06E-04	[9.72E-05 ,	1.16E-04]	2.00E-04	[1.75E-04 ,	2.28E-04]	2.39E-04	[2.30E-04 ,	2.47E-04]
IE	3.39E-07	[1.83E-07 ,	5.52E-07]	2.24E-05	[1.88E-05 ,	2.61E-05]	2.34E-04	[1.56E-04 ,	3.26E-04]	3.95E-05	[3.04E-05 ,	4.91E-05]
IT	6.56E-06	[5.22E-06 ,	8.06E-06]	5.81E-05	[5.51E-05 ,	6.11E-05]	4.52E-05	[3.52E-05 ,	5.72E-05]	1.34E-06	[8.14E-07 ,	2.00E-06]
LT	2.08E-05	[7.6E-06 ,	3.5E-05]	1.2E-04	[1.0E-04 ,	1.4E-04]	1.3E-04	[4.6E-05 ,	2.5E-04]	3.8E-02	[3.7E-02 ,	3.9E-02]
LU	3.11E-05	[5.6E-06 ,	5.2E-05]	2.8E-05	[1.2E-05 ,	5.0E-05]	4.0E-04	[1.9E-04 ,	6.9E-04]	6.8E-04	[5.5E-04 ,	8.1E-04]
LV	2.85E-06	[9.1E-08 ,	1.2E-05]	7.5E-05	[6.0E-05 ,	9.2E-05]	7.5E-05	[3.2E-06 ,	3.0E-04]	9.4E-05	[8.8E-05 ,	1.0E-04]
NL	5.37E-06	[3.7E-06 ,	6.7E-06]	2.0E-05	[1.8E-05 ,	2.2E-05]	1.5E-04	[1.1E-04 ,	1.8E-04]	2.9E-05	[2.6E-05 ,	3.1E-05]
PL	2.01E-05	[1.7E-05 ,	2.0E-05]	5.8E-05	[5.4E-05 ,	6.3E-05]	2.1E-05	[1.5E-05 ,	2.9E-05]	3.9E-05	[3.4E-05 ,	4.5E-05]
PT	8.37E-06	[6.8E+03 ,	9.8E-06]	8.2E-06	[7.3E-06 ,	9.1E-06]	5.4E-06	[2.9E-07 ,	1.8E-05]	2.3E-06	[7.4E-07 ,	4.9E-06]
SE	1.46E-04	[2.4E-05 ,	2.4E-04]	7.7E-05	[6.3E-05 ,	9.1E-05]	5.0E-03	[3.1E-03 ,	7.1E-03]	3.8E-04	[2.3E-04 ,	5.4E-04]
SI	6.52E-06	[2.2E-07 ,	2.9E-05]	1.3E-04	[1.1E-04 ,	1.5E-04]	8.8E-05	[5.5E-05 ,	1.3E-04]	2.3E-04	[2.1E-04 ,	2.4E-04]
SK	1.50E-07	[5.5E-09 ,	6.6E-07]	3.8E-04	[3.5E-04 ,	4.1E-04]	6.4E-04	[5.0E-04 ,	8.0E-04]	1.1E-03	[1.0E-03 ,	1.1E-03]
UK	1.71E-06	[1.1E-06 ,	2.0E-06]	4.4E-05	[4.0E-05 ,	4.8E-05]	1.2E-03	[1.1E-03 ,	1.4E-03]	4.8E-04	[4.7E-04 ,	4.9E-04]

Table 14. Estimated values for *qi*, *Salmonella* subtype-dependent factor (mean and 95% Credibility Interval).

Serovar	*qi*	95% CI
***S*. Enteritidis**	1[(a)]	
***S*. Agona**	0.0527	[0.0488 , 0.0569]
***S*. Anatum**	0.0252	[0.0223 , 0.0283]
***S*. Bovismorbificans**	0.1854	[0.1690 , 0.2034]
***S*. Brænderup**	0.1386	[0.1223 , 0.1567]
***S*. Brandenburg**	0.1096	[0.1009 , 0.1190]
***S*. Bredeney**	0.0170	[0.0151 , 0.0191]
***S*. Derby**	0.0197	[0.0186 , 0.0201]
***S*. Hadar**	0.0734	[0.0670 , 0.0806]
***S*. Heidelberg**	0.1163	[0.0960 , 0.1401]
***S*. Infantis**	0.1223	[0.1167 , 0.1281]
***S*. Kentucky**	1.9980	[1.7970 , 2.2130]
***S*. Kottbus**	0.0143	[0.0124 , 0.0164]
***S*. Livingstone**	0.0595	[0.0540 , 0.0653]
***S*. London**	0.0826	[0.0751 , 0.0908]
***S*. Mbandaka**	0.0473	[0.0425 , 0.0523]
***S*. Montevideo**	0.1124	[0.1044 , 0.1210]
***S*. Newport**	0.2476	[0.2320 , 0.2645]
***S*. Rissen**	0.0302	[0.0268 , 0.0340]
***S*. Saintpaul**	0.0600	[0.0538 , 0.0671]
***S*. Typhimurium**	0.2153	[0.2054 , 0.2264]
***S*. Virchow**	0.2469	[0.2320 , 0.2625]

(a) The q value for *S*. Enteritidis is fixed to 1, and the other serovars are calculated relatively to it.

6.4. Model goodness-of-fit

The predictive ability of the model was assessed by estimating the ratio between the observed *Salmonella* cases (sporadic human cases reported in each country) and the number of cases predicted by the model and attributable to sources in each country. A ratio of one reflects a perfect model fit, whereas a ratio higher than 1 means that the model tends to underestimate the number of cases, and an estimate below 1 refers to an overestimation. Results of the test showed that the model fit was satisfactory for the vast majority of the countries (Figure 19). Poor fit was observed for countries with poor data availability or quality, e.g. Cyprus and Luxembourg. The need for complete data is clearly represented in the goodness-of-fit graph, as

the countries which are further from the 1.0 axis are the ones which did not report outbreak data (Cyprus, Greece, Italy, Luxembourg) or travel information (Belgium, Spain, France). The low fit for Cyprus is also likely to be a reflection of the consumption data, as it had to have a large amount of data estimated, and was the only low PSI value when comparing the estimated data with the "real" FAO profile, as seen in the data management chapter. Luxembourg, on the other hand, was characterized earlier by its small animal sample sizes, which could be a reflection of a small production, but also of a non-representative sample which may have had a negative impact on the results.

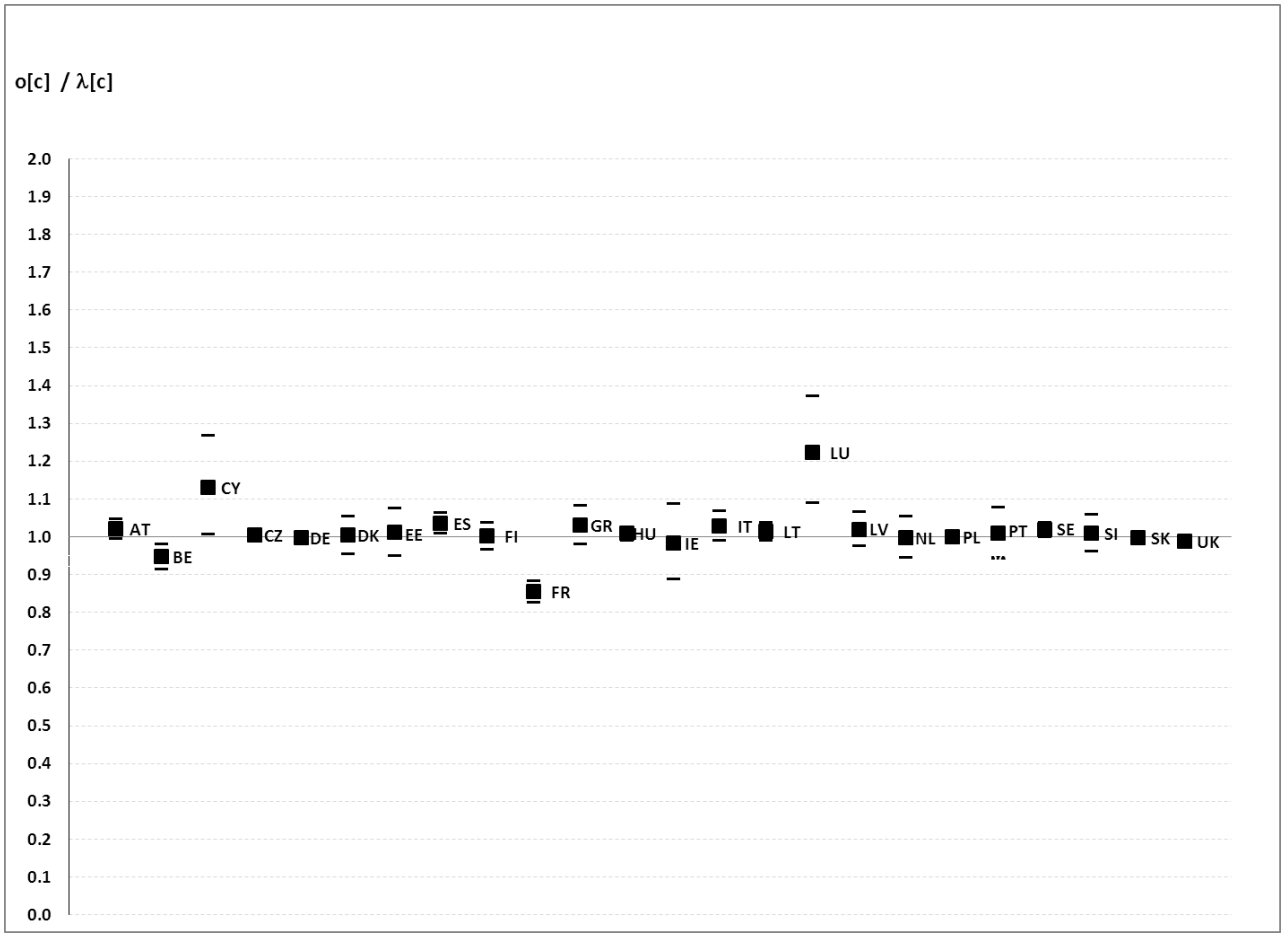


Figure 19: Ratio between observed and predicted cases of *Salmonella* in 24 EU Members States with 95% inter-percentile range. A ratio of 1 represents a perfect model fit.

6.5. Comparison of Danish attribution estimates between the EU model and the Danish model (Manuscript III)

The objective of this section is to compare the results of the Danish *Salmonella* source account as published in the Annual Report on Zoonoses in Denmark each year with the source attribution results obtained for Denmark in the EU model, as an attempt to validate both models and propose improvements to both approaches. This comparison proved important because the two models aimed at attributing human salmonellosis in Denmark to food/animal sources, but differed in several aspects, particularly regarding the data used for prevalence and serovar distributions (Table 15). The impact of these differences, in particular in the attribution estimates from imported food (i.e. of foreign origin) was investigated.

Although the mathematical approach is the same in the two models, differences arise when it comes to data used. As an example, the EU model included data from 24 countries and these data included only serotyping information and varied in representativeness and quality; on the contrary, the Danish model makes use of data with higher discriminatory power, i.e. with a better resolution of subtyping. On the other hand, the surrogate consumption data (i.e. production and trade data) used in the EU model are more detailed.

Regarding the prevalence data (p), the EU model uses data from the EFSA BSs, whereas the Danish model uses a combination of risk-based data from the Case-by-case Risk Assessment Program (CBC) and from national monitoring programs. The CBC started in 2007 and collects samples from batches of Danish and imported pork, beef, chicken, turkey and duck, which are then tested for *Salmonella*. Apart from the CBC, individual retail samples are also collected from domestic and imported ducks and turkeys. Results are recorded by country of origin, but the prevalence in imported sources enters the model as an overall percentage of positive samples by type of imported meat. *Salmonella* data from domestic sources are also collected from pork and beef carcass swabs taken at slaughter and from broiler and layer flocks of *Gallus gallus*.

For the amount of a food source available for consumption (m), the estimated amount of each source imported from each country (in tonnes, EUROSTAT data) were used in the EU model; the Danish model uses the total imported amount of a source (in tonnes), without specifying the origin. In practical terms, this means that the EU model ultimately works in four dimensions, since the country of origin of the food and to which the *Salmonella* prevalences apply can differ from the country where the human cases were reported, and cases are consequently attributed to the countries from where the food originated. This combination of data types used for *p* and *m* also means that, while the Danish model uses *m* only for weighting the sources among themselves, in the EU model the prevalence in a country exporting a large amount of a food source to Denmark will have a higher impact on the results, when compared to countries exporting smaller amounts.

A summary comparison between the two methods is shown in Table 15, and those differences will support the presentation of the results.

Table 15. Data-related features of the Danish source account model and the EU source attribution model.

	Danish model	EU model	Comment
Human data included	Data from 2007, 2008 and 2009	Aggregated case-based data from 2007 to 2009	No year-specific inferences are possible in the EU-model.
Source of human data	Statens Serum Institute (SSI).	ECDC / EFSA after reporting from countries.	Locally produced and reported data have fewer opportunities for information loss from the point of collection to the point of storage. Danish data in the EU model was reported to ECDC by SSI.
Travel information	Cases with missing information modeled according to the probabilities observed in the ones with information, resulting in 31% travel and 69% domestic.	Cases with missing information assumed to be domestic, resulting in 18% travel and 82% domestic.	The Danish model assumes that the follows the same distribution as the information provided. The EU model assumes that cases not referred specifically as travel-related are domestic, mainly because some countries had 0% travel information, and it was not possible to estimate the proportion of travelers. This assumption is likely to result in an underestimation of travel cases in the EU model, as some of the not-specified cases would be travel-related.
Subtyping information	Most isolates serotyped, *S*. Enteritidis and *S*. Typhimurium phage typed and *S*. Typhimurium tested for susceptibility to nine antimicrobials.	Serovar level used. The serovar distribution of cases and samples with missing serovar information were modeled based on observed distributions in the relevant datasets, resulting in a larger uncertainty regarding the true serovar distribution. This was particularly the case in the human datasets.	Higher level of detailing attributes cases more specifically to the right sources, but also leave a relatively higher proportion of cases with "unknown source", as the model requires a "perfect match" between subtypes in humans and animal reservoirs. On the other hand, in the model with less subtype detailing, cases could be misplaced, as the same serovar can be present in different sources, and the source with higher prevalence will "draw" more cases.
Food/animal sources included and origin of *Salmonella* prevalence data,	Domestic: pork, beef, broilers, layers and ducks (from national surveillance programs). Imported: pork, beef, chicken, ducks and turkeys (from the case-by-case risk assessment program and retail monitoring).	Pigs, broilers, turkeys and layers (from EFSA baseline studies or EU-harmonized surveillance). Differentiation between imported and domestic based on the EUROSTAT production and trade data (see below).	The fewer the number of sources included in the model, the more likely it is for cases to be attributed to a wrong source. As an example, beef is absent from the EU model; however, *S*. Typhimurium is an important serovar in both cattle and pigs, and it is likely that some *S*. Typhimurium cases which were caused be beef are attributed to pigs in this model. Another expected resulting difference of the two approaches is that in the Danish modelimported eggs are not included, as theyare generally considered to be of low importance, as they are mainly used for heat-treated products by the industry and there is consequently no monitoring of imported shell eggs. O; in the EU model, they enter as a source, where the impact is determined by the imported amount and the prevalence in the country of origin.
Consumption data	Domestic and imported amounts of each source available in the country, with no differentiation between countries of origin of imported food.	Estimated from production, exports and imports reported to EUROSTAT. Specific amounts originating from each exporting country available.	The use of trade data, allows discrimination among foods originating from different countries, particularly when country-specific prevalences are available from the BS studies. The use of these data bring along some biases and assumptions, as described in the methods section.
Model dimensions	Subtype (serovar, phage type, resistance pattern), source and year	Serovar, source, country of human case reporting and country of origin of food	The "country of origin of food" dimension allows the attribution of cases to the country of origin of the sources.

6.4.1. Overview of results from the two models

A total of 7,433 human cases of *Salmonella* were reported in Denmark in the period from 2007 to 2009. Table 16 shows the number of reported cases attributed to animal reservoirs, international travel and outbreaks in each year in the Danish model, as well as the sum of the three years. The most important sources of salmonellosis in this period were pork (7.9% domestic and 1.4% imported, resulting in 9.3%), table eggs (7.5%) and broilers (4.7%), of which imported chicken (3.8%) was the largest part. Around 30% of the total cases reported were estimated to have been acquired abroad, and 16.7% of all cases could not be attributed to any source.

Table 16 DK model: Estimated percentage of *Salmonella* cases attributed to food/animal sources, international travel, outbreaks with source unknown and unknown sources, 2007-2009, Denmark (mean and 95% Credibility Interval).

Source	2007	2008	2009	2007-2009
Broilers	0.3 (0.0-1.0)	1.3 (0.7-3.6)	0.7 (0.1-1.8)	0.9
Imported chicken	1.4 (0.4-2.8)	5.2 (3.3-6.8)	3.7 (2.1-5.3)	3.8
Pork	7.6 (6.0-9.3)	8.8 (7.6-10.0)	6.5 (3.6-9.7)	7.9
Imported pork	2.0 (1.0-3.1)	0.5 (0.3-1.9)	1.3 (0.2-2.8)	1.4
Turkeys	-	0	-	0
Imported turkey	2.0 (0.5-3.5)	2.4 (0.2-4.1)	0.7 (0.1-1.8)	1.9
Table eggs	12.3 (11.5-13.2)	3.2 (2.5-3.9)	11.0 (8.9-13.2)	7.5
Beef	0.2 (0.1-0.3)	0.7 (0.4-1.0)	0.7 (0.1-1.6)	0.6
Imported beef	3.1 (2.2-4.0)	0.3 (0.1-0.7)	1.2 (0.6-1.8)	1.3
Ducks	0.3 (0.0-0.9)	1.0 (0.1-2.7)		0.6
Imported duck	1.4 (0.5-2.3)			0.4
Travel	32.2 (30.4-31.4)	23.3 (23.1-23.6)	46.3 (44.4-48.2)	30.6
Unknown source	17.7 (15.1-19.8)	13.1 (11.3-15.0)	23.4 (20.0-26.8)	16.7
Outbreaks, unknown source	20.9	39.6	4.4	26.4
TOTAL	2,129	3,656	1,647	7,433

In the EU model, 7,461 cases of salmonellosis were reported in Denmark in the period of 2007 to 2009. After adjusting for underreporting (see section 5.1.2.1.), this resulted in 26,331 cases (Table 17), with turkeys as the most important food source of sporadic cases (19.6%), followed by pigs (18.0%), layers (10.1%) and broilers (3.5%). When including also non-food sources, most cases were attributed to international travel (23.7%). Cases that could not be attributed to any source corresponded to 18.3%, and outbreaks with unknown source had 6.8% of cases.

Table 17. EU model: Estimated percentage of *Salmonella* cases attributed to animal reservoirs, international travel, outbreaks with source unknown and unknown sources, 2007-2009, Denmark.

Source	Total source percentage[(a)]	Percentage by origin[(b)]
Broilers	3.5 (0.5-12,5)	0
Imported broilers		3.5
Pigs	18.0 (3.2-61.4)	14.7
Imported pigs		3.3
Turkeys	19.6 (2.9-69.2)	0
Imported turkeys		19.6
Layers	10.1 (2.3-33.1)	1.2
Imported eggs		8.9
Travel	23.7 (3.6-83.0)	23.7
Unknown source	18.3 (2.8-64.0)	18.3
Outbreaks, unknown source	6.8	6.8
Total	26,330	26,330

(a) Results of the EU model. See Appendix A.
(b) Total source percentage divided based on country "originating" Danish cases. For percentages "originated" from all MSs included in the model in the four sources, see Appendix F.

Figure 20 shows the attributable percentages to categories divided by domestic or imported source. The category "others" contains sources present in the Danish model but not in the EU model (e.g. beef and ducks).

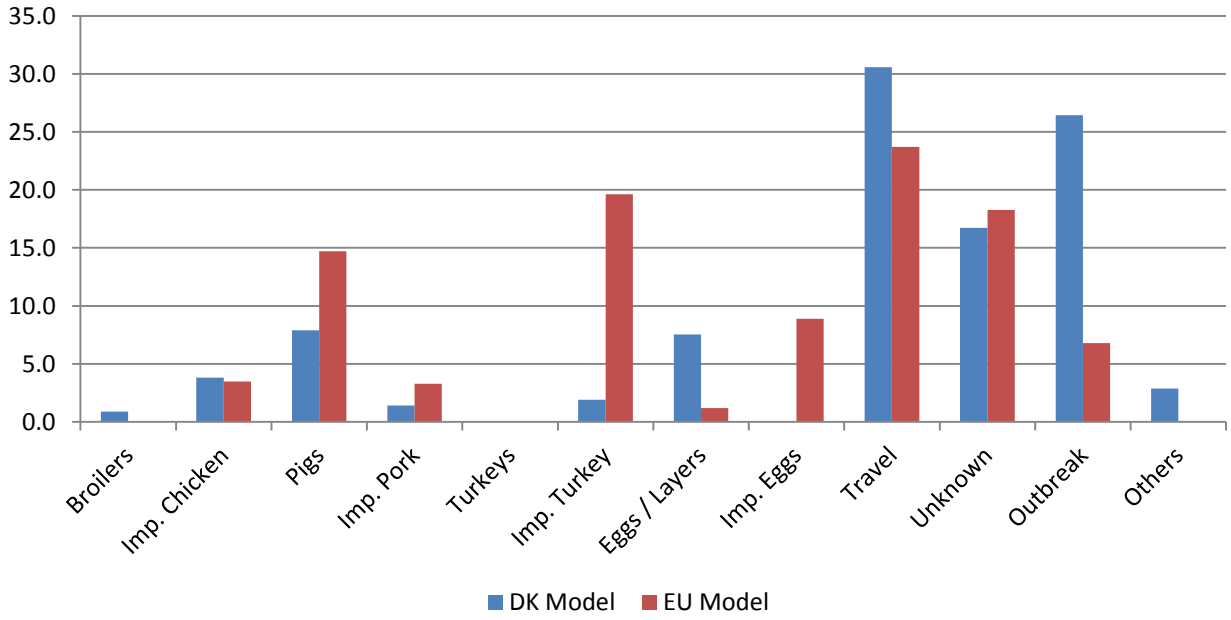


Figure 20. Attributable fractions of *Salmonella* cases to domestic and imported animal sources in Denmark in the Danish model and in the EU model, 2007-2009.

The differences in the proportion of cases attributed to the animal sources between the two models result in great part from the differences in the data, categories and model assumptions already described. The largest discrepancy is observed in the results for turkeys. Concerning the other sources, although individual attributed fractions are different, the order of priority indicated by the results is similar, having pigs as the most important, followed by layers and then broilers. The four animal sources and the travel-related cases will be addressed in separate in the next section. To help comparing the results of the two models, Tables 18, 20, 21 and 22 show the overall *Salmonella* prevalence and the data used for *m* in the countries estimated to be responsible for the cases attributed to the sources.

The amount of cases attributed to the "unknown" category is affected by the different number of sources in the two models and by the higher level of subtyping detail in the DK model (phage types, AMR profile). The more sources that are included in the model, the more likely it is for a case to be attributed to the right source. At the same time, a more discriminatory level of subtyping makes it more difficult to "match" cases to the right sources, resulting in a larger number of cases being directed to "unknown". In the EU-model, due to the less detailed subtyping information, some cases may be wrongly attributed to one of the four sources. This is particularly true for the more common serovars, such as *S.* Typhimurium. This is also one of the reasons for the animal sources in general receiving a larger proportion of cases when compared to the Danish model.

The greater part of the difference in the proportion of cases attributed to outbreaks lies in the fact that the results of the EU model were adjusted for underreporting, with a UF of 4.4. It was assumed that all outbreak-related cases were properly reported, and so outbreak cases were not multiplied by the UF. This changes the balance between the proportion of cases attributed to outbreaks and to the other categories, when comparing the two models. If the results of the Danish model are multiplied by 4.4 and the proportions are re-calculated, cases belonging to outbreaks with an unknown source change from 26.4% to 7.6% of the total, which is reasonably more similar to the 6.8% estimated by the EU model.

6.4.2. International travel

The Danish model estimates the proportion of cases with no travel information that are travel-related, assuming that these should follow the same proportions as the ones for which that information is available; as a consequence, the total cases attributed to travel includes reported and estimated cases. Because no travel information was available for some countries, cases with no travel information were by default considered domestic in the EU model, whereas a part of these were attributed to travel in the Danish model. In the EU model dataset, *S.* Enteritidis corresponds to 46% of travel cases in Denmark. Due to control activities conducted in the country during the last decade, the prevalence of that serovar in food-animals is in general low, particularly when compared to other MSs. For that reason, a proportion of these cases were attributed to imported foods, particularly imported eggs and broilers, indicating that the serovar distribution in imported food is comparable to the serovar distribution in travelers.

6.4.3. Turkeys

Both models agree that all cases attributed to this reservoir are related to imported turkey. However, the proportion of cases attributed in the EU model is over 10 times the proportion in the Danish model. In Table 18, it is evident that the total amount of turkey meat imported by Denmark as considered in the Danish model was smaller than in the EU model, resulting in a smaller parcel of cases weighted to this reservoir. The data used in the Danish model also shows that the CBC tested samples from the four countries responsible for 88% of cases in the EU model (Germany, France, Italy and Poland), which were also the

main exporting countries in the period, according to EUROSTAT. This shows that the CBC seems to have a good sensitivity, testing samples from countries which both models agree are important. At the same time, in the EU model, the use of trade data to estimate the available amount for consumption makes it possible to use specific prevalences applied to specific imported amounts, giving a weighted importance to high-prevalence countries exporting large amounts of food to Denmark.

Table 18. Comparison of the overall Salmonella prevalence and amount available for consumption in the two models and percentage of the number of cases reported in Denmark attributed to turkeys by the EU model

Exporting country	**% of Danish cases attributed to turkeys**	**EU model**		**DK model**	
		prevalence	***m***	**prevalence**	***m***
BE	0.1	10.8	80	N/A	N/A
DE	24.3	7.3	14,102	15.6	N/A
ES	1.5	39.1	222	N/A	N/A
FR	23.6	9.6	6,021	8.7	N/A
HU	7.2	62.5	782	N/A	N/A
IE	0.1	22.7	228	N/A	N/A
IT	9.3	20.2	2,968	46.4	N/A
LT	0.1	4.4	77	N/A	N/A
NL	0.9	9.0	512	N/A	N/A
PL	30.9	17.7	5,695	39.4	N/A
UK	1.9	25.5	783	N/A	N/A
Total	100.0	-	31,470	18.75	23,687

Looking more closely at Poland, Germany and France, responsible for almost 80% of cases, Table 19 shows the prevalence of *S.* Saintpaul, one of the most important turkey serovars (as seen in section 5.1.2.2.1.), and of *S.* Enteritidis and *S.* Typhimurium, the two most important overall. The adjusted number of human cases of each of those serovars in Denmark for the period was, respectively, 352, 7,044 and 6,310 cases.

Table 19. Prevalence of selected serovars in the four animal sources included in the EU model in Poland (PL), Germany (DE) and France (FR)

Country	**Serovar**	**Prevalence (p)**			
		Broilers	**Pigs**	**Turkeys**	**Layers**
PL	*S.* Saintpaul	0.24	0.00	6.77	0.01
	S. Enteritidis	7.16	2.47	0.93	10.11
	S. Typhimurium	2.39	1.19	3.04	0.52
DE	*S.* Saintpaul	0.00	0.00	1.57	0.00
	S. Enteritidis	0.00	0.40	0.14	2.50
	S. Typhimurium	4.86	9.19	1.82	0.39
FR	*S.* Saintpaul	0.00	0.09	0.61	0.03
	S. Enteritidis	0.24	0.18	1.17	2.18
	S. Typhimurium	0.00	7.83	1.47	1.31

As observed, the prevalence of *S.* Saintpaul in Polish turkeys is 4.3 times the observed value for Germany, and over 10 times the French prevalence, while it is almost absent in the other three sources in the three countries. But more importantly, Poland is the only country in which the prevalence of *S.* Typhimurium is higher in turkeys than in any other source, suggesting that, if phage typing data or more detailed subtyping methods were available, part of the cases attributed to turkeys might have been attributed to other sources (for instance, pigs). Given the large number of observed *S.* Typhimurium cases (6,310), this results in a large difference, suggesting that the EU model is likely to have overestimated the importance of this source, particularly the contribution from Poland, pointed as the main contributor of turkey-originated cases (31%).

6.4.3. Broilers

When comparing results for broilers (Figure 20), one immediately visible difference is the absence of cases attributed to domestic broilers in the EU model. This is readily explained by the different data used; in the BS, the prevalence of *Salmonella* in broiler carcasses in Denmark was zero, while the surveillance and monitoring data used in the Danish model had 2.1% positive samples. The parcel attributed to (imported) broilers in the Danish model is also larger than in the EU model. This can probably be explained by the lower level of subtyping detail in the EU model. As prevalences for *S.* Enteritidis were consistently higher in layers than in broilers, without better discriminatory features a parcel of *S.* Enteritidis broiler cases are likely to have been attributed to layers. As 27% of sporadic human cases in Denmark (7,044 out of 26,330) were caused by this serovar, this parcel corresponds to a large proportion of total cases. In addition, the Danish model includes data from meat imported from non-EU countries, such as Brazil, Chile and Argentina. The EU model does not take those countries into consideration, which could result in non-EU broiler cases being "forced" into the available countries in this model (Table 20). As an extra note, according to the EUROSTAT data the UK exports a large amount of broiler meat to Denmark, but no positive samples from this country were reported in the CBC.

Table 20. Comparison of the overall Salmonella prevalence and amount available for consumption in the two models and percentage of the number of cases reported in Denmark attributed to broilers by the EU model

Exporting country	% of Danish cases attributed to broilers	EU model		DK model	
		prevalence	m	prevalence	m
AR[(a)]	-	N/A	N/A	6.7	N/A
BE	9.4	20.3	7,335	8.4	N/A
BR[(a)]	-	N/A	N/A	12.0	N/A
CZ	0.4	5.5	416	N/A	N/A
DE	16.5	17.6	26,935	9.2	N/A
ES	7.2	14.9	2,418	N/A	N/A
FR	2.1	7.6	8,644	10.0	N/A
GR	0.1	14.8	70	N/A	N/A
HU	9.7	85.7	1,442	N/A	N/A
IE	2.0	9.9	218	N/A	N/A
IT	0.4	16.8	894	N/A	N/A
LT	0.5	6.9	2,299	6.6	N/A
LV	0.1	4.9	27	N/A	N/A
NL	2.6	10.0	23,773	42.9	N/A
PL	30.2	25.5	6,597	3.6	N/A
PT	7.0	11.2	1,633	N/A	N/A
SE	0.4	0.2	71,499	4.9	N/A
SI	1.0	1.7	3,426	N/A	N/A
SK	0.3	21.6	51	N/A	N/A
UK	9.9	3.5	8,287	N/A	N/A
Total	100.0	-	165,964	8.6	93,191

(a) Non-EU countries from where Denmark has imported chicken meat

6.4.4. Pigs

Pigs present a particular situation, in which cases attributed to imported sources only represent a small fraction of the total. In the EU model, 81.5% of cases attributed to this source are estimated to come from Denmark (Table 21), and these results are consistent with the fact that 84.9% of the pork fraction in the Danish model were attributed to domestic pork (Table 16). The consistency between the two models is further demonstrated as the overall prevalences in the two datasets are reasonably similar in the two largest contributors besides Denmark (Germany and Spain) (Table 21). This suggests that the difference in attributable fractions is more likely due to the differences in the total imported amount and the difference in discriminatory power, which in this case plays an important role: *S.* Typhimurium, the most important pig serovar, is one of the main serovars in all other sources, being also responsible for the second largest amount of human cases. Without better discriminatory power, a large parcel *S.* Typhimurium cases is attributed to this source in the EU model, which corresponds to a large parcel of total cases. In the Danish model, phage typing data allows better differentiation among sources, resulting in less cases being directed to this source.

Table 21. Comparison of the overall Salmonella prevalence and amount available for consumption in the two models and percentage of the number of cases reported in Denmark attributed to pigs by the EU model

Exporting country	% of Danish cases attributed to pigs	EU model		DK model	
		prevalence	m	prevalence	m
BE	0.6	13.0	11,840	N/A	N/A
DE	6.8	12.7	123,767	10.0	N/A
DK	81.5	8.0	3,013,472	3.1	N/A
ES	5.9	30.7	62,648	33.3	N/A
FR	1.1	18.5	22,896	29.6	N/A
HU	0.3	11.6	3,611	N/A	N/A
IE	0.6	15.4	10,592	N/A	N/A
IT	0.1	16.4	4,355	N/A	N/A
NL	1.4	8.5	46,638	16.7	N/A
PL	0.7	0.7	11,069	N/A	N/A
UK	1.0	1.0	12,969	31.8	N/A
Total	100.0	-	3,323,857	11.9	230,440

6.4.5. Layers / Eggs

The resulting parcels attributed to layers/eggs are very similar in the two models (especially if looking at 2007 and 2009 in Table 19, as an outbreak of unknown source in 2008 reduced the relative importance of other sources). However, in the EU model 88% of cases attributed to layers come from imported eggs (Table 22), while the Danish model only considers domestic eggs, so the similarities do not come from consistency between models, as happened for pigs. *Salmonella* contribution from imported eggs is not considered important in Denmark, as it is believed that most of the imported eggs are not sold as shell eggs, but instead used in heat-treated products. Whether this assumption holds is not known.

As happened with turkeys, the specific serovar prevalences in Table 23 provide an indication of the reasons for the discrepancy between models. The prevalence of *S.* Enteritidis, the main serovar in layers, is 50 times higher in Poland than in Denmark, which also has lower prevalences of the other two important layer serovars, *S.* Typhimurium and *S.* Infantis. Also, specific phage types of *S.* Enteritidis, like PT 21, PT 4 and PT 6, are most frequently related to travel in Denmark. This information cannot be taken into consideration in the EU model, as phage type information is not available. The model, therefore, tends to allocate those cases to countries from which Denmark imports eggs with high *S.* Enteritidis prevalence. As mentioned earlier, *S.* Enteritidis and *S.* Typhimurium are the most observed serovars overall, so parcels of cases attributed to any of them result in high attributed proportions at country level.

Table 22. Comparison of the overall Salmonella prevalence and amount available for consumption in the two models and percentage of the number of cases reported in Denmark attributed to layers by the EU model

Exporting country	% of cases Danish cases attributed to layers	EU model		DK model	
		prevalence	m	prevalence	m
AT	0.1	2.5	341	N/A	N/A
BE	0.9	11.7	1,060	N/A	N/A
CZ	0.3	8.9	167	N/A	N/A
DE	4.8	3.5	8,999	N/A	N/A
DK	11.9	0.6	200,645	5.42	82,594
ES	5.8	44.5	1,080	N/A	N/A
FR	0.1	6.1	121	N/A	N/A
LV	2.3	20.3	829	N/A	N/A
NL	4.1	2.6	7,595	N/A	N/A
PL	69.6	12.5	32,450	N/A	N/A
SE	0.1	0.7	2,763	N/A	N/A
Total	100.0	-	256,050	N/A	82,594

Table 23. Prevalence of selected serovars in the four animal sources included in the EU model in Denmark (DK), Poland (PL), Spain (ES), Germany (DE) and the Netherlands (NL)

Country	Serovar	Prevalence (p)			
		Broilers	Pigs	Turkeys	Layers
DK	*S.* Enteritidis	0.00	0.00	0.00	0.20
	S. Typhimurium	0.00	4.84	0.00	0.20
	S. Infantis	0.00	0.91	0.00	0.00
PL	*S.* Enteritidis	7.16	2.47	0.93	10.11
	S. Typhimurium	2.39	1.19	3.04	0.52
	S. Infantis	6.21	0.26	1.24	0.90
ES	*S.* Enteritidis	5.40	0.37	0.58	24.03
	S. Typhimurium	1.29	15.95	1.62	2.82
	S. Infantis	0.51	0.00	0.00	6.43
DE	*S.* Enteritidis	0.00	0.40	0.14	2.50
	S. Typhimurium	4.86	9.19	1.82	0.39
	S. Infantis	1.39	0.32	0.00	0.17
NL	*S.* Enteritidis	0.00	0.00	0.00	2.60
	S. Typhimurium	0.48	5.54	0.70	0.04
	S. Infantis	0.72	0.10	0.00	0.00

Part II:

An alternative approach for source attribution in countries with missing data

7. SOURCE ATTRIBUTION BY EXPERT ELICITATION BASED ON CLUSTER ANALYSIS OF NON-HEALTH VARIABLES

7.1. Background

The Hald model (Hald et al., 2004) and its variations (Pires and Hald, 2010; Toyofuku et al., 2011; Whalström et al., 2011; Guo et al., 2011) are among the most widespread microbial subtyping-based methods for source attribution of *Salmonella*. As shown in section 5 of this thesis, these methods require a large amount of good-quality data, which are available from the Danish surveillance system and, up to a point, from datasets maintained by EUROSTAT, studies published by EFSA and national harmonized surveillance systems. These data requirements reduce the applicability of the models, as nationally representative and internationally-comparable studies on the baseline prevalence of *Salmonella* in production animals, human data from laboratory-based integrated surveillance systems and trade data are only available from a limited number of countries. However, in places where those data are not present, other types of data may be available from public sources such as FAO, UNDP or published papers which could, when analyzed by an appropriate group of experts, provide a way to approximate the results of source attribution models. Expert elicitations can be used to estimate the proportion of disease attributed to food sources (Batz et al., 2005). These opinions serve to supplement data collection, in a similar way as meta-analyses and systematic reviews supplement primary research. Expert elicitation can be used to assess probability or risk ranking based on personal experiences and/or relevant information, but where direct measurements are not possible (Hoffmann et al., 2007a). This has been done for source attribution in several countries, such as the United States (Hoffmann et al., 2007a), the Netherlands (Havelaar et al., 2008), New Zealand (Lake et al., 2010) and Canada (Davidson et al., 2011).

7.2. Objective

The objective of this part of the thesis was to propose and evaluate an alternative approach for source attribution based on expert elicitation, using non-health indicators as information to estimate results for countries where the data on *Salmonella* required for the Bayesian model are not available. The chosen approach was to provide a panel of experts with sets of countries grouped in accordance wih social, economic, environmental and dietary characteristics, as well as with sets of countries grouped according to the results of the EU source attribution model presented in Part I of this thesis. The experts were then asked to provide source attribution estimates for countries not included in the EU model. The main assumption of this approach is that the parcels of human salmonellosis attributed to different animal sources indirectly reflect the social, economic, environmental and dietary characteristics of a country. The knowledge of experts may then be used to relate A) countries for which attribution estimates have been obtained by the microbial subtyping approach in Part I with B) countries for which the only data available are those that might be indirectly associated with the sources of human salmonellosis, e.g. the data collected through FAO, UNDP and climate data.

7.3. Methods

7.3.1. Choosing non-*Salmonella* variables

Although representative surveys on the presence of *Salmonella* in livestock are not available from most countries in the world, it is known that the introduction and transmission of pathogens in animals of the food chain depend largely on the type of production system, including type of housing and animal density, which are, by turn, influenced by local environmental patterns (average temperatures, rainfall, etc)

(FAO/OIE, 2009). Production systems are also affected by livestock production/demand relationships, which are largely influenced by social and economic characteristics of the country (FAO, 1995). As an example, the level of economic growth of a country is a key factor determining the demand for livestock products, as it brings an increase in disposable income. This is normally followed by an increase in migration to urban areas, reducing the size of the population involved in the primary production, at the same time as it increases the number of consumers. In countries where this has happened, production systems had to be adapted and become more intensive (FAO, 1995). So, from a wider point of view, social and economic characteristcs of a population can provide indirect information about the presence of pathogens, including *Salmonela*, in production animals in a country.

When it comes to the presence and survival of *Salmonella* during processing and consumption, climatic conditions can influence the viability, stability and growth rates in the environment, as high temperatures increase the replication cycles of most food- and waterborne bacterial pathogens (Semenza et al., 2012). *Salmonella*, in particular, has optimal growth temperature around 37 degrees Celsius, and in the absence of other preservation methods, temperature is expected to influence its growth at various points in the food chain, particularly inside the household (Adams and Moss, 1995; Kovats et al., 2004). Furthermore, a linear association between temperature and the number of cases of salmonellosis reported nationally has been observed in the Netherlands, England and Wales, Switzerland, Spain and the Czech Republic, normally with the effect being observed one week after the temperature increase (Kovats et al., 2004).

Still concerning *Salmonella* at the point of consumption (i.e. in the household), food-related behaviors are complex and determined by the interplay of many factors, among which education, income, ethnicity and food availability (De Irala-Estevez, 2000). Such behaviors include food item ingestion patterns, as well as food preparation habits, which are key in preventing or allowing the survival of a pathogen in the household. Consumption habits are also expected to have an influence on which subtypes will be normally found in a population, as reservoir-specific subtypes will be absent if those reservoirs are not used as food, or their presence will vary depending on how each of those reservoirs are prepared.

Finally, countries have different foodborne disease surveillance systems, resulting in large variations in the level of reporting and the surveillance and monitoring activities performed. Factors that may contribute to the variations include economic development, access to health care, public health infrastructure and demographic features (rural/urban, literacy, age, religion, food preferences), among others (WHO, 2002). Figure 21 illustrates factors that affect the presence, survival and transmission of *Salmonella* (in blue) in different steps of the path (in gray) from the production of a contaminated animal reservoir at farm level until the reporting of a human case of foodborne salmonellosis. The figure also shows non-health variables that may reflect or influence those factors and which are publically available from the United Nations Development Program (UNDP, in pink), the United Nations Food and Agriculture Organization (FAO, in green), EFSA (in red) or published studies (in purple).

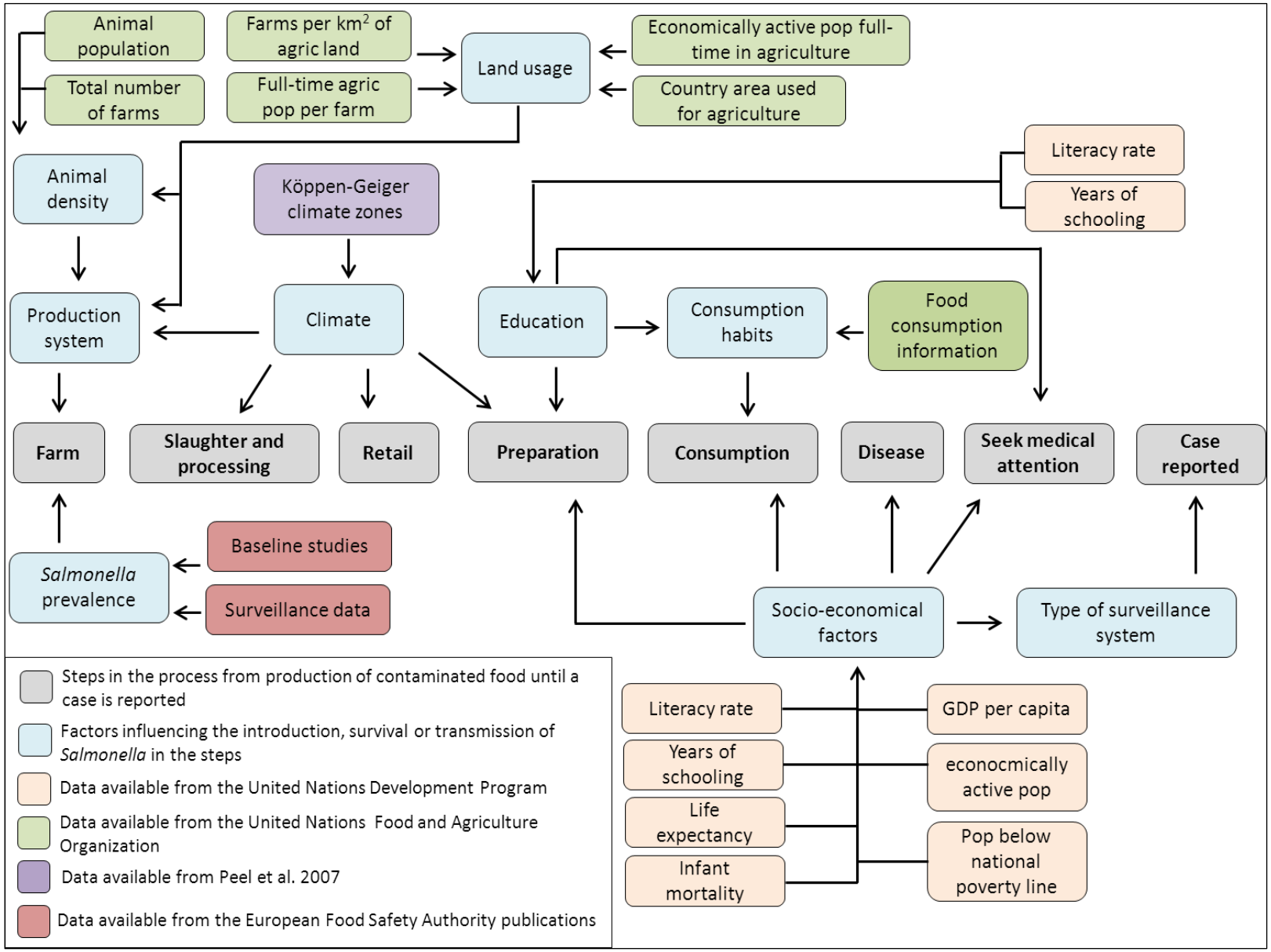


Figure 21. Non-health variables with potential to be used as tools for source attribution of *Salmonella* using expert elicitation.

Based on the above, the chosen variables were:

- relative proportions of consumption of eggs, poultry meat, pork, beef, sheep and goat meat, fish, seafood, raw animal fats and "other meats";
- consumption of poultry meat (g/person/day);
- consumption of pork (g/person/day);
- consumption of eggs (g/person/day);
- gross domestic product (GDP) per capita in U.S. dollars;
- percentage of the population which is economically active;
- percentage of the population below the national poverty line;
- literacy rate (%);
- mean years of schooling among adults;
- life expectancy in years;
- mortality under five years of age (per 1000 births);
- percentage of country territory used for agriculture;
- percentage of economically active population working full-time in agriculture;
- number of farms per square kilometer of agricultural land;
- number of individuals employed full-time in agriculture per farm unit;
- chickens per farm;
- pigs per farm;
- turkeys per farm;
- climate information

7.3.2. Variables derived from results of the EU model

Besides the non-*Salmonella* information, the countries which were included in the EU model were also clustered according to the results presented in Part I. Variables derived form those results were:

- *Salmonella* incidence attributable to all sources;
- *Salmonella* incidence attributable to broilers;
- *Salmonella* incidence attributable to pigs;
- *Salmonella* incidence attributable to turkeys;
- *Salmonella* incidence attributable to layers;
- attributable fraction of human *Salmonella* cases to all sources combined;

- relative proportion of reported *S*. Enteritidis, *S*. Typhimurium and "Other serovars" in humans;
- relative proportion of *S*. Enteritidis, *S*. Typhimurium and "Other serovars" in broilers;
- relative proportion of *S*. Enteritidis, *S*. Typhimurium and "Other serovars" in pigs;
- relative proportion of *S*. Enteritidis, *S*. Typhimurium and "Other serovars" in turkeys;
- relative proportion of *S*. Enteritidis, *S*. Typhimurium and "Other serovars" in layers;

7.3.3. Grouping of countries using cluster analysis

Cluster analysis was used to identify similarities among countries in relation to the chosen variables. Hierarchical cluster analysis starts with all observations in a dataset belonging to the same cluster, and systematically creates new clusters, by separating observations which are more similar among themselves than to the others in relation to a set of variables. The procedure can be performed progressively until each observation constitutes its own cluster.

In this study, an "average subject" from each cluster was chosen as the centroid to be compared with other clusters, and the squared Euclidean distance between observations within the same cluster was used as similarity measure. The more similar the subjects, the smaller the distance between them (and consequently, the smaller the squared Euclidean distance), the same principle applying to less-similar subjects and larger distances. Variables measured in different scales which were used in the same set were standardized to fit a distribution with mean=0 and standard deviation=1. It is necessary to standardize the values before running the analysis, otherwise variables that differ thousands of units from each other (e.g. country territory in squared kilometers) will drive the cluster construction, annulling the influence of variables that vary in a smaller scale (e.g. percentages).

The resulting process can be plotted as a dendrogram (or "tree") with the distance between clusters on the vertical axis. Although the whole hierarchical structure can be visualized in this way, the best cluster solution was chosen for each set of variables to be presented. This choice was based on an evaluation of the clustering process using a) the root-mean-square deviation (RMSSTD) of each new cluster formed, b) the semipartial R-squared (SPR), c) the R-squared (RS) and d) the distance between two clusters (CD). These measures provide a statistical reference to evaluate the homogeneity of a new cluster formed and the heterogeneity among the current group of clusters in each step, indicating the more "natural" number of clusters for a given set of observations. Cluster analyses methodology and goodness-of-fit measures are detailed in Sharma (1996).

7.3.4. Expert elicitation

Countries included in the study were Austria, Bulgaria, Czech Republic, Cyprus, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, the Netherlands, Norway, Poland, Portugal, Slovenia, Slovakia, Spain, Sweden and the United Kingdom. Experts were asked to provide estimates of the percentage of cases of human salmonellosis in Bulgaria, Norway and Romania which originated from the broiler, pig, laying hen and turkey reservoir, travel and "unknown or other sources", according to how countries grouped based on the "indirect *Salmonella* indicators" and on the results of the EU microbial subtyping model. The point of attribution was the animal reservoir, as consequence of the method used and in line with the original study in Part I of this thesis. The steps for an expert elicitation process as described by Hoffmann et al. (2007b) were followed.

7.3.4.1. Determining the size of the expert panel

A panel of 12 experts was selected. Because the aim of the study was to pilot the approach, the number of experts was found to be sufficient. Also, when the research question is narrowly defined in terms of discipline or expertise, the elicitation can rely on a small panel (Cooke and Goossens, 2004).

7.3.4.2. Choosing the model of analysis and/or aggregation

No weighting was used in this pilot, as the panel was considered small and relatively homogeneous in area of expertise. A set of validity checks based on the protocol by Hoffmann et al. (2007b) was used to evaluate the quality and homogeneity of the information provided by the panel. These included 1) variability in expert judgment; 2) the level of agreement between the experts' assessments and prior estimates based on primary data; 3) individual experts' uncertainty about their own assessments; 4) variability in individual experts' uncertainty about their own best estimates. Although the statistical analysis of those check points was not possible in the present study because the original protocol was developed for large panels, the concepts behind them were applied when evaluating the experts' guesses. Variability in judgment, agreement with other experts and individual uncertainty were evaluated visually using box plots, after analyzing the estimates as described in section 7.3.4.7. Agreement of panel and individual assessments based on primary data was a challenge, as the objective was to obtain estimates for which there are no other results. For that reason, the Czech Republic was included as one of the countries under study and used as a validation reference, with expert guesses for this country being compared with the estimates from the EU model. This country was arbitrarily chosen because members of the panel could have already read the original results published as part of the contract with EFSA (Pires et al., 2011a), and among the first choices for the panel there were no experts originating from the Czech Republic, reducing the chances of them remembering and looking up those results. No other countries were added as validation reference, as the number of countries available for clustering was considered already too small, and this would reduce this number even further.

Experts' estimates of attributable proportions were fit as theoretical Beta Pert distributions, using the point estimates as mode and the limits of the uncertainty range as maximum and minimum. This distribution was chosen as it was specifically designed for modeling expert estimates, and it is more sensitive to the most likely value than to the minimum and maximum values, making it a better choice than, for instance, the Triangle distribution, which is the other commonly used distribution for this kind of assessment (Vose Software, 2007. Reference Number: M-M0361-A). These distributions were then used to generate aggregated panel estimates through the following process: experts were numbered 1 to 7, and a random number generator was set up to draw 10,000 values between 1 and 7. With each drawn number, a random sample was taken from the distribution derived from the guess of the corresponding expert, creating a joint distribution which included the uncertainty of each one. The mean of the resulting distribution for each source was used as the final guess from the panel. Original estimates and intervals were stored using Microsoft Excel version 14.0.61112.5000 (32-bit) ©Microsoft Corporation. Simulations were run in ModelRisk Standard version 4.3.1.1. © Vose Software 2011.

7.3.4.3. Choosing the mode of elicitation

Typically, expert elicitations use group interaction or one-on-one interviews to elicit expert judgment (Cook, 1991 *apud* Hoffmann et al., 2007b). One-on-one assessments, by turn, are generally conducted through in-depth interviews or a written elicitation instrument (Hoffmann et al., 2007a). In this study, individual assessments using a written instrument were chosen as the mode of elicitation.

7.3.4.4. Developing the elicitation survey instrument

The prepared materials provided the experts with the results of the EU model, the data used to obtain those results and other data intended to design "country profiles", which allowed comparisons among countries to include those without enough data for the microbial subtyping approach. The task was presented in an instrument divided in two parts, namely the instructions/questionnaire and the information sheets.

7.3.4.4.1. Instructions and questions

Each expert received, along with the information sheets, a set of instructions and study objectives, as well as a written elicitation instrument under the form of a blank table with the countries and sources, as seen in Figure 22. The material also included five questions intended to collect the experts' opinion on the validity and utility of the data used and the method in general. The panel was asked to analyse how countries without attribution results related to countries for which those results were available, regarding a set of economic, social, agricultural and climate data, and also information on *Salmonella* incidences and serovar proportions in humans and animals from the original EU model dataset, when those were available. Based on that analysis, they were asked to provide their estimates for the fraction of cases attributable to broilers, pigs, layers, turkeys, travel and unknown sources in the Czech Republic, Bulgaria, Romania and Norway. Intervals containing the minimum and maximum values for their estimates were also requested, as those intervals allow experts to express their uncertainty and also reduce the respondent's fatigue (Havelaar et al., 2007). Finally, the experts were asked to answer five questions evaluating the usefulness and applicability of the approach. In order to evaluate the attribution results, estimates for the Czech Republic were also asked from the panel, so values can be compared with the ones obtained by the microbial subtyping approach for that country. The full elicitation instrument is presented in Appendix G.

2 - Fill in the attributable fractions (%) you estimate for each source in Bulgaria, the Czech Republic, Norway and Romania, adding a range for your answer. If not enough information was provided for an estimate, write "NP" (for "not possible") in the corresponding field.

Source	Czech Republic			Bulgaria		
	%	Range		%	Range	
Broilers						
Pigs						
Turkeys						
Layers						
Travel						
Unknown / other reservoirs						

Source	Romania			Norway		
	%	Range		%	Interval range	
Broilers						
Pigs						
Turkeys						
Layers						
Travel						
Unknown / other reservoirs						

Figure 22: Tables used to collect source attribution estimates from the expert panel.

7.3.4.4.2. Information sheets

Information sheets were prepared using the results of the cluster analyses. In total, 27 information sheets were developed, of which the first three were instructions on how to read and interpret the dendrograms and the tables. The remaining sheets can be classified in nine major groups, which contain sheets with resulting clusters for one variable, as well as multi-variable sheets, maps, graphs or excel spreadsheets. A Pearson's correlation analysis was run among the variables included in each group, to assess if any of them would have a significant influence on the others. A correlation coefficient lower than 80% was found for all analyses, and so it was considered that the variables could be included in the planned clustering groups. The main groups and their composition were as it follows:

1. *Source attribution outcomes*. This group contains results of the source attribution approach based on microbial subtyping in 24 EU countries presented in section 6, and should be used as a reference to estimate attributable fractions to animal reservoirs in countries without attribution studies. *Salmonella* data used for humans, broilers, turkeys, layers and pigs were the same as described in section 5.1.2.2. Human incidences were corrected for underreporting with the use of underreporting factors (Havelaar, 2012). The incidences refer to a period of three years (2007-2009), and are presented in cases/100,000. Although the percentage of travel-related cases is shown in the bar graph, it was not included in the cluster analysis, as differences between countries were too large and would obscure the importance of the contribution of animal reservoirs. Sheets included were:
    a. *Salmonella* incidence attributable to all sources (overview table);
    b. *Salmonella* incidence attributable to broilers (table + dendrogram);
    c. *Salmonella* incidence attributable to pigs (table + dendrogram);
    d. *Salmonella* incidence attributable to turkeys (table + dendrogram);
    e. *Salmonella* incidence attributable to layers (table + dendrogram);
    f. attributable fraction of human *Salmonella* cases to all sources combined (overview table);
    g. attributable fraction of human *Salmonella* cases to all sources combined (dendrogram);
    h. cumulative attributable fractions bar graph;
2. *Relative proportions of S.* Enteritids, *S.* Typhimurium *and "Other serovars" in humans and animal sources in each country* (EU model dataset) (5 sheets):
    a. relative proportion of reported *S*. Enteritidis, *S*. Typhimurium and "Other serovars" in humans;
    b. relative proportion of *S*. Enteritidis, *S*. Typhimurium and "Other serovars" in broilers;
    c. relative proportion of *S*. Enteritidis, *S*. Typhimurium and "Other serovars" in pigs;
    d. relative proportion of *S*. Enteritidis, *S*. Typhimurium and "Other serovars" in turkeys;
    e. relative proportion of *S*. Enteritidis, *S*. Typhimurium and "Other serovars" in layers;
3. *Food consumption information* (FAO, 2003) (4 sheets):
    a. relative proportions of consumption of eggs, poultry meat, pork, beef, sheep and goat meat, fish, seafood, raw animal fats and "other meats";
    b. consumption of poultry meat (g/person/day);

c. consumption of pork (g/person/day);

d. consumption of eggs (g/person/day);

4. *Economic indicators* (UNDP, 2011). This group contains one sheet in which countries were clustered according to three variables:

   a. gross domestic product (GDP) per capita in U.S. dollars;

   b. percentage of the population which is economically active;

   c. percentage of the population below the national poverty line;

5. *Non-economic human development indicators* (UNDP, 2011). This group contains one sheet in which countries were clustered according to four variables:

   a. literacy rate (%);

   b. mean years of schooling among adults;

   c. life expectancy in years;

   d. mortality under five years of age (per 1000 births);

6. *Agriculture and land usage characteristics* (FAO, 2011). This group contains one sheet in which countries were clustered according to four variables:

   a. percentage of country territory used for agriculture;

   b. percentage of economically active population working full-time in agriculture;

   c. number of farms per square kilometer of agricultural land;

   d. number of individuals employed full time in agriculture per farm unit;

7. *Density of animal production* (FAO, 2011). This group contains one sheet in which countries were clustered according to three variables together:

   a. chickens per farm;

   b. pigs per farm;

   c. turkeys per farm;

8. *Climate data*. This sheet contains a map of Europe showing Köppen-Geiger climate zones as updated by Peel et al. (2007), as well as a table extracted from the original article with a description of Köppen climate symbols and defining criteria. No cluster analysis was performed, as national borders and climate zones do not always coincide.

9. *Cluster results summary* (Excel file). This group contains "country X country" matrices based on the best solution for each set of variables, showing:

   a. in which information sheets each two countries belonged in the same cluster;

   b. the probability that two countries belonged in the same cluster in the study, calculated by dividing the number of times they were clustered by the number of times they could be clustered, as not every country was present in every analysis.

The origin of the data for each analysis and the construction of composed variables, as well as the demonstration and information sheets, are presented in Appendices H and I, respectively. The full set of information sheets is available upon request.

7.3.4.5. Identifying the expert pool

Twelve experts were selected from three institutions involved in public health and foodborne diseases, based on availability and on expertise in risk modeling and epidemiology. The panel was balanced in terms of gender, with five males and seven females. Six experts were Danish, three from the Netherlands, one from Sweden, one from Portugal and one from Brazil. The research institutions were located in Denmark, the Netherlands and the United States. Each member was approached personally or by email and invited to participate.

7.3.4.6. Administering the elicitation survey

Upon acceptance to take part in the study, an email containing the study objectives and general approach was sent with the materials for the elicitation as an attachment.

7.3.4.7. Analyzing the survey results

Experts' guesses of the attributable fraction of salmonellosis cases to each source were plotted to compare the most likely values and uncertainty ranges among experts, and those who were consistently different from the rest of the panel were excluded. Individual Czech Republic estimates and aggregated panel estimates for the Czech Republic were also compared against reference values from the EU model to evaluate how accurate they were, and a Proportionaliry Similarity Index was calculated following the same methodology described for the trade data, to compare the order of priority among sources between the results of the panel and those obtained by the microbial subtyping approach.

7.4. Results

Estimates were received from seven out of 12 experts. One expert did not consider the information provided enough to give estimates for Romania. Results referring to the Czech Republic will be presented first in separate, as they are used as reference to evaluate the quality of the elicitation.

7.4.1. Czech Republic

Although the values were different from the EU model (Tables 24 and 25), all experts maintained the same order of priority among animal sources as in the reference, namely layers, pigs, broilers and turkeys. Travel, however, was identified as a less important source than broilers by experts 1, 5, and 6, which changes the dynamic when considering all types of known sources. The "Unknown/other reservoirs" category shows the largest differences when comparing with the reference results, as in practical terms it became an uncertainty depository, where parcels not distributed among known sources were allocated.

The relative estimated proportions among animal sources with the uncertainty range (minimum and maximum possible values) for each expert are better visualized in Figures 23a to 23d. Concerning the evaluation of the elicitation quality (section 7.3.3.1), the additional visualization of the plots for Bulgaria, Norway and Romania (Appendix K) shows that experts were reasonably consistent about the sources of which they were more or less certain in all four countries, and concordance among experts was also considered good, except for experts 2 and 7. Due to the frequent discordance between Expert 7 and the panel

or the reference values, and to the large individual uncertainty of Expert 2, they were excluded, as it was considered that their guesses reduced the quality of the joint estimates further from the reference values. This is confirmed in Figure 24, where the joint estimates for the full panel (seven experts) and the filtered panel (five experts) for the animal sources are plotted along with the reference values.

Table 24. Estimated fractions of cases attributed to different sources in the Czech Republic from the EU model.

EU model	Mean	2.50%	97.50%
Broilers	0.1	0	0.2
Pigs	10.9	10.2	11.5
Turkeys	1.7	1.4	2.1
Layers	83.9	82.8	85
Travel	1.7	-	-
Unknown/Other reservoirs	0.8	0.0	1.8

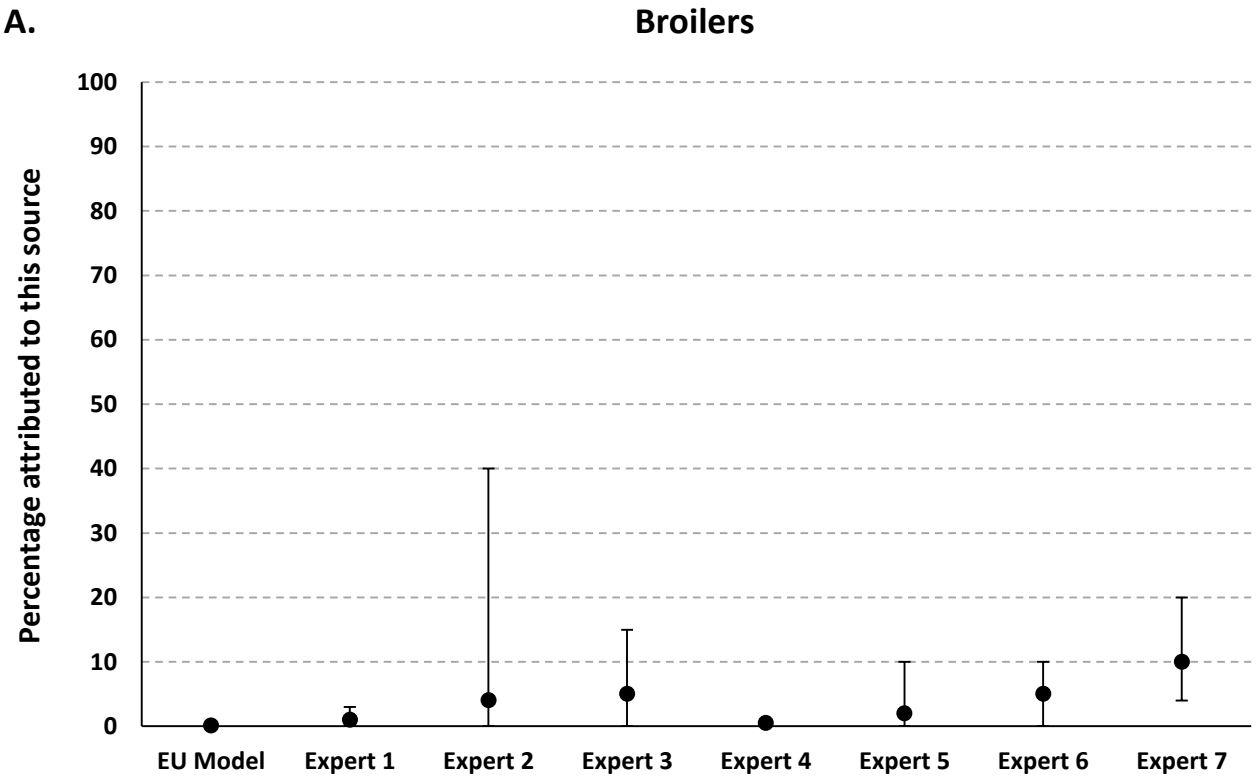
Table 25. Estimated proportion of salmonellosis cases estimated by each expert with minimum and maximum possible values uncertainty range) for the Czech Republic.

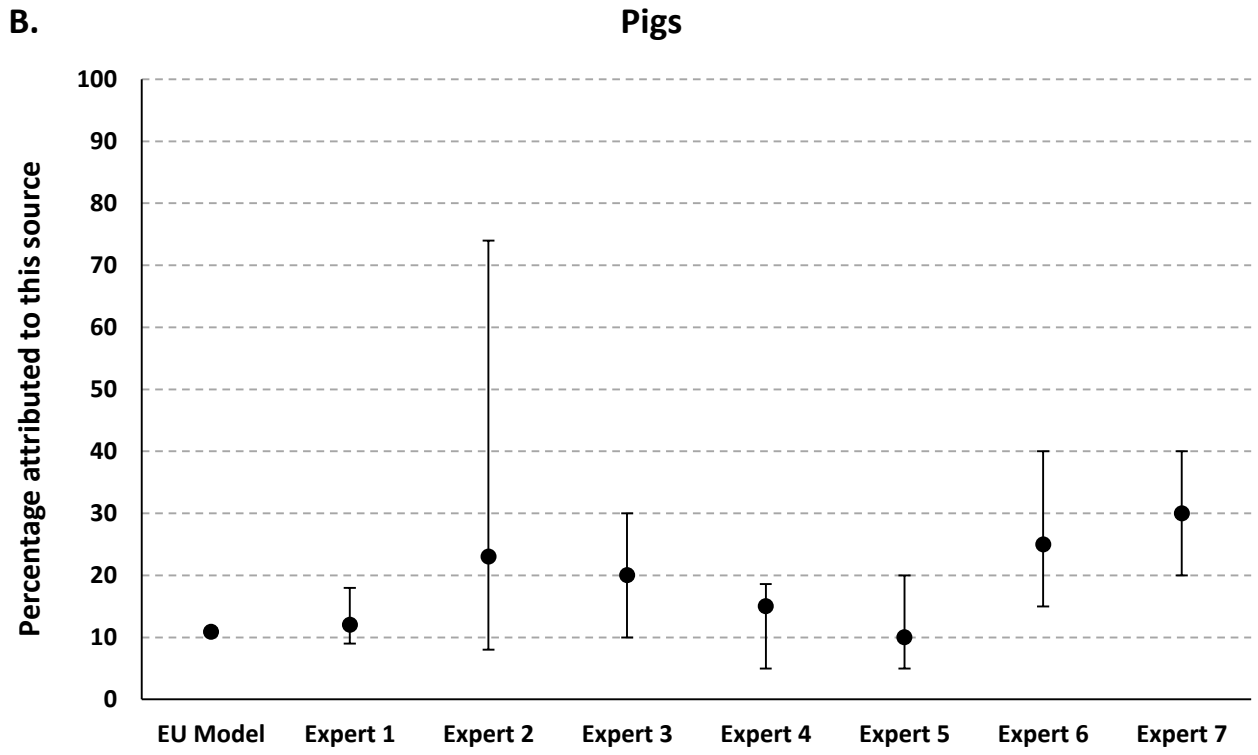
Respondant	Estimate	Minimum	Maximum
Expert1			
Broilers	1	0.1	3
Pigs	12	9	18
Turkeys	1	0.1	4
Layers	70	5.5	85
Travel	2	0.1	5
Unknown/Other reservoirs	14	5	20
Expert2			
Broilers	4	0	40
Pigs	23	8	74
Turkeys	5	0	15
Layers	47	2	83
Travel	5	0	30
Unknown/Other reservoirs	16	4	38
Expert3			
Broilers	5	0	15
Pigs	20	10	30
Turkeys	5	0	15
Layers	50	30	70
Travel	5	0	15
Unknown/Other reservoirs	15	5	30
Expert4			
Broilers	0.5	0	1
Pigs	15	5	18.6
Turkeys	3	2	3.1
Layers	55	25	76.1
Travel	5	5	5
Unknown/Other reservoirs	21.5	20	23
Expert5			
Broilers	2	0	10
Pigs	10	5	20
Turkeys	1	0	5
Layers	75	60	85
Travel	1	0	3
Unknown/Other reservoirs	11	0	15
Expert6			
Broilers	5	0	10
Pigs	25	15	40
Turkeys	5	2	10
Layers	50	30	65
Travel	1	0	3
Unknown/Other reservoirs	14	5	40
Expert7			
Broilers	10	4	20
Pigs	30	20	40
Turkeys	5	2	10
Layers	35	20	50
Travel	5	2	10
Unknown/Other reservoirs	15	8	20

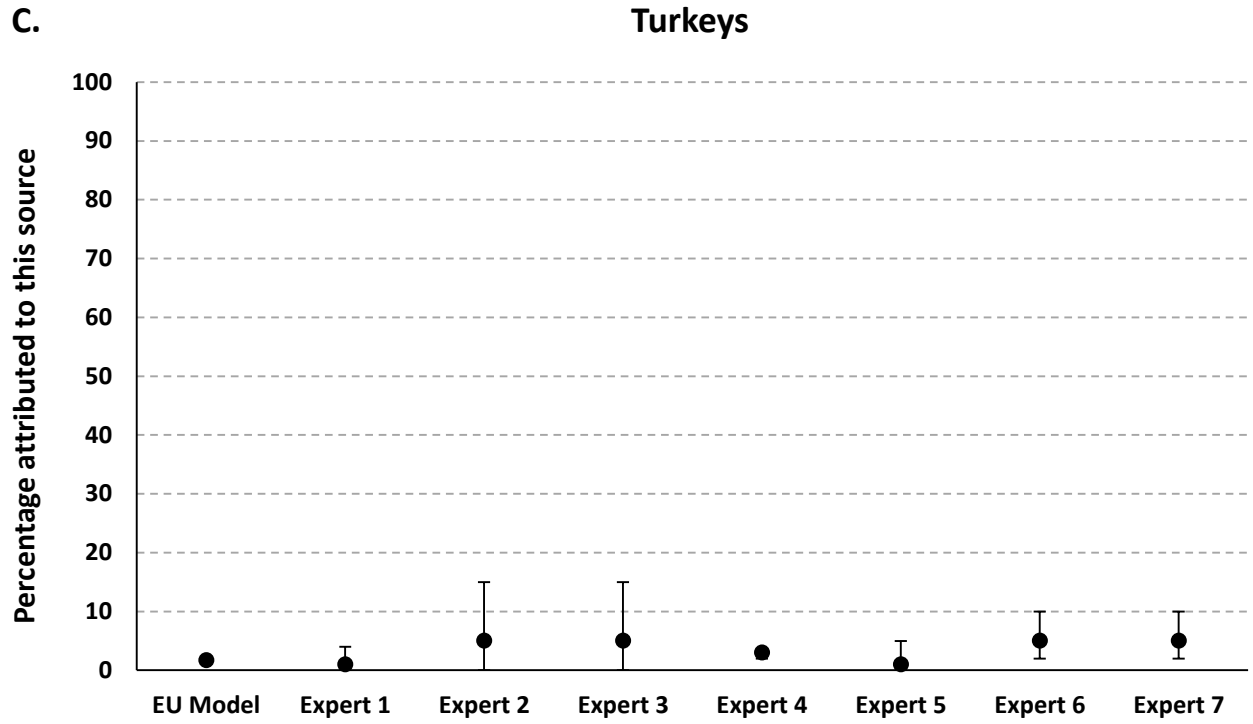
The joint estimates for the Czech Republic are shown in Table 26. As also found for the individual estimates, the proportions attributed to each source by the filtered panel were numerically different from the ones estimated by the EU model. When comparing the proportions attributed among sources by the two approaches, 73.8% similarity was observed for the results including the "Unknown" category, and 83.5% without it, showing that the elicited estimates for this country are in good agreement with the ones obtained by the microbial subtyping approach.

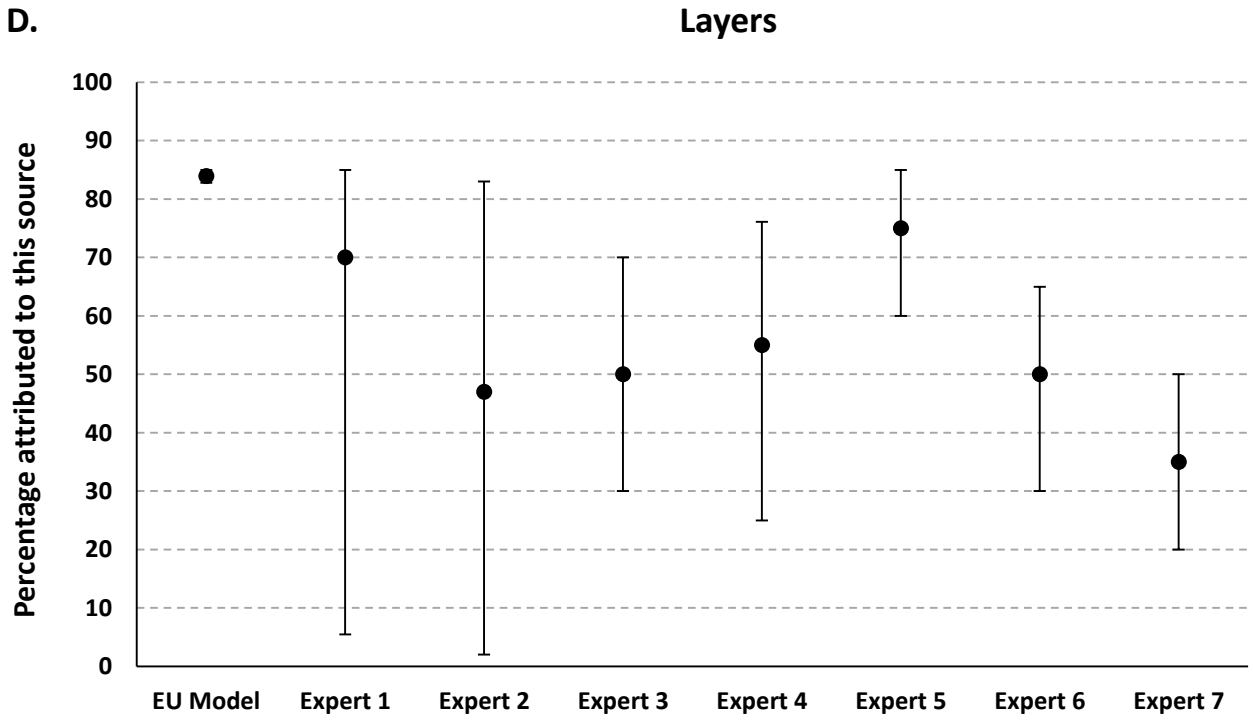
Table 26: Joint panel estimates for all sources in the Czech Republic.

Source	**Filtered panel**			**Full panel**		
	Mean	**95% CI**		**Mean**	**95% CI**	
Broilers	3.1	0.3	8.2	5.1	0.4	14.3
Pigs	16.6	8.5	29.0	20.4	9.2	36.0
Turkeys	3.4	0.5	8.1	4.0	0.6	8.7
Layers	57.8	38.5	79.0	52.6	29.0	77.9
Travel	3.1	0.5	7.9	4.1	0.6	10.5
Unknown/Other reservoirs	15.4	7.3	22.9	15.8	7.9	24.3


A.
Broilers
Percentage attributed to this source
100
90
80
70
60
50
40
30
20
10
0
EU Model
Expert 1
Expert 2
Expert 3
Expert 4
Expert 5
Expert 6
Expert 7


B.
Pigs
Percentage attributed to this source
100
90
80
70
60
50
40
30
20
10
0
EU Model
Expert 1
Expert 2
Expert 3
Expert 4
Expert 5
Expert 6
Expert 7





Figures 23a to 23d. Individual expert guesses (most likely value, minimum and maximum) plotted against results from the EU model (mean, standard deviation) in the Czech Republic.

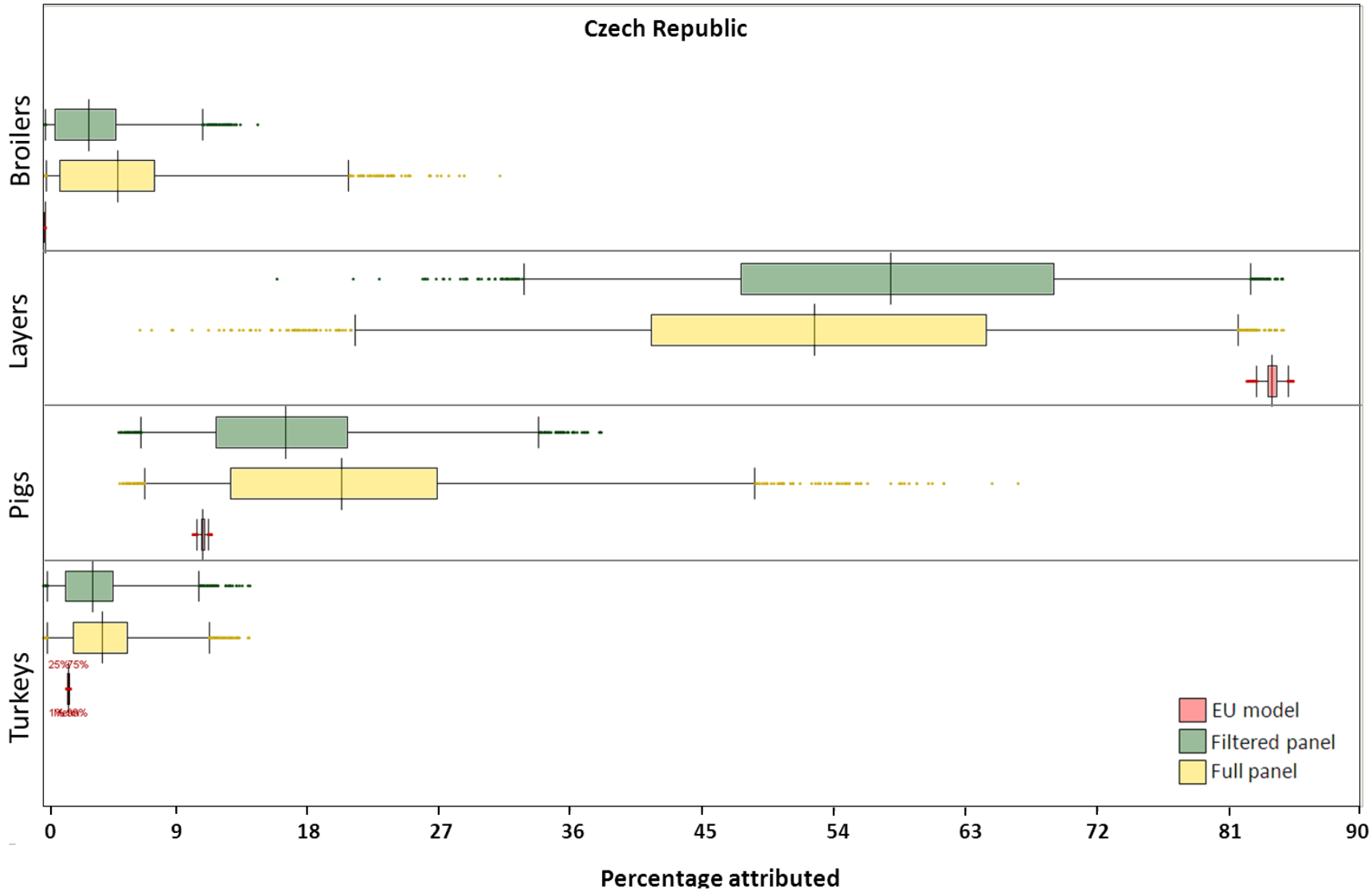


Figure 24. Joint-estimate distributions from the full and filtered panels plotted against results from the EU model in the Czech Republic.

7.4.2. Bulgaria, Norway and Romania

Individual estimates received and the corresponding plots for Bulgaria, Norway and Romania are shown in Appendices J and K. It was assumed that the same quality of estimates observed for the Czech Republic, as well as the filtering of Experts 2 and 7, could be extrapolated to the other countries. This was done because the behavior of the experts appeared to be reasonably consistent throughout the elicitation, as seen on Appendix K. Final panel estimates for Bulgaria, Norway and Romania are shown in Table 27, and the box plots for those values are shown in Appendices L1, L2 and L3.

Table 27: Joint panel estimates for all sources in Bulgaria, Norway and Romania.

Country	Source	Filtered panel			Full panel		
		Mean	95% CI		Mean	95% CI	
Bulgaria	Broilers	2.6	0.5	7.9	3.5	0.6	9.6
	Pigs	10.9	4.6	19.9	14.1	4.9	35.8
	Turkeys	1.1	0.0	2.7	2.6	0.1	8.9
	Layers	73.8	51.8	84.9	66.7	35.0	84.1
	Travel	1.7	0.2	3.7	3.6	0.3	10.2
	Unknown/Other reservoirs	9.4	3.9	15.9	12.2	4.4	26.2
Norway	Broilers	0.9	0.3	1.8	1.8	0.3	6.3
	Pigs	4.6	2.3	7.5	6.2	2.5	11.8
	Turkeys	1.6	0.5	2.7	6.1	0.5	37.0
	Layers	3.0	1.3	5.8	4.1	1.4	8.7
	Travel	80.1	74.7	85.3	72.2	23.4	84.8
	Unknown/Other reservoirs	10.6	6.4	15.2	11.3	4.0	22.4
Romania	Broilers	2.9	0.5	8.0	5.4	0.6	18.8
	Pigs	15.6	4.6	27.5	15.5	0.0	34.2
	Turkeys	2.2	0.1	7.7	2.6	0.0	9.4
	Layers	64.8	44.4	81.5	51.7	0.0	81.0
	Travel	1.8	0.4	3.9	2.4	0.0	9.8
	Unknown/Other reservoirs	12.2	4.1	21.9	11.6	0.0	23.8

7.4.3. Evaluation of the method by the experts

To better evaluate the method, experts were asked about how they approached the given task and what kind of logical thinking guided their working process. Three basic "attacking the problem" tactics were registered by the experts:

1) thinking of the process in a risk assessment or statistical source attribution point of view, keeping in mind an equation of a known model and trying to use information from surrogate countries to approximate each component of the equation and estimate results;
2) trying to determine one "surrogate country" for each of the ones under study and adapting their attribution results according to differences observed, for example, in food consumption patterns;
3) trying to determine one "surrogate country" for each of the ones under study and use the same attribution results;

Depending on the information available for each country, the same expert used one or more of the tactics above. Only one expert mentioned the use of the summary tables to look for the surrogate countries, directly using the pairs with the highest probability of grouping. The others preferred to look at specific variables which they considered key to the process. There was a consensus among members of the panel that the social, economic and climatic variables were not used at any moment. One member suggested that this may have happened due to their extensive knowledge of the area under study, replacing the information provided with their own concepts. The information sheets mentioned as crucial for the process were the source attribution results, the *Salmonella* incidences, food consumption and the serovar profiles, in which case the main objective of the approach was not attained, as the non-health information available worldwide was not considered useful, and the main instruments in the process were the information more limitedly available. There was also consensus from the panel on the limitations of the applicability of the method, as most suggest that using a regression model to obtain significant parameters that can be used as input in an equation to obtain estimates, or developing a method for using surrogate data in the traditional microbial subtyping models are more sensible ways to go. Regarding the specific attribution results, the five experts composing the filtered panel commented that the "unknown category" received the remaining proportions to sum 100% after the guesses for the other sources were made, rendering the results for this category not useful.

8. DISCUSSION

This thesis presented two multi-country approaches for source attribution of human salmonellosis. A microbial subtyping approach for source attribution was applied to data from 24 EU Member States and required extensive data management and validation. The second approach, which used clustering techniques and expert elicitation, used the results of the first approach as input and extrapolated to European countries with insufficient *Salmonella* data. This section discusses the methods individually, as well as the overall results of conducted studies.

8.1. The EU model

Results of the EU-wide microbial subtyping model suggest that layers were the most important reservoir of salmonellosis in the EU, being responsible for over 40% of *Salmonella* infections at the time of the study and found as the main source in 13 out of 24 MSs. Pigs were estimated to be the second largest contributor at EU level and the main one in eight MSs, while turkeys were revealed as particularly important only in Denmark and broilers in Portugal. These results are in general in line with previous source attribution estimates obtained at regional and national level in European countries.

An EFSA scientific opinion that used a comparative analysis and interpretation of *Salmonella* serovar occurrence in animals and humans has reached similar conclusions, having also estimated layers and pigs to be the first and second source of salmonellosis (EFSA, 2010e). Given the scope of the study, a quantitative assessment was only available for slaughter pigs, attributing 10 to 20% of cases to this source, which is less than the 31% attributed by the EU model. However, these figures were described as "guesstimates", and the report recommended the development of a microbial subtyping model for the EU, which was expected to provide more reliable results. A source attribution model using outbreak data applied by Pires et al. (2011a) to attribute salmonellosis in the EU in the period from 2007 to 2009 used different categories for the food items included, but also estimated eggs as the main source of human salmonellosis in the EU, followed by pork.

Another EU attribution study, which has been conducted after the work of this thesis was finalized for the setting of targets for *Salmonella* control in the turkey production in the EU (Hald et al., 2012), used the same mathematical principle as the one presented here, but updated and slightly different datasets. This model (which we refer to as the TT-SAM model) used the same human data as our model, but included 25 MSs and used 2010 data for all sources except pigs, for which also the BS data were used. Results pointed in a different direction when compared to ours and to the other two studies discussed, having estimated pigs as the main reservoir of salmonellosis (56.8%), followed by layers (17.0%), broilers (10.6%) and turkeys (2.6%). Neverteless, it is important to note that the total number of reported human cases in EU has decreased in the period from 2007 to 2010 (EFSA, 2011a; EFSA, 2012a), from approximately 154,000 cases (8.9 million after adjusting for underreporting) to 99,000 (5.41 after adjusting for underreporting). This reduction is largely explained by a reduction in the number of *S.* Enteritidis cases, which are particularly associated with shell-egg production (Hald et al., 2004; Pires et al., 2009; EFSA, 2012a, Pires et al., 2011a), thus affecting the overall source attribution estimates and the relative contribution of food sources. The prevalence of *S.* Enteritidis in laying hens has decreased significantly in the same time period (EFSA, 2012a), which has been linked to the EU harmonised monitoring put in place in 2008 and the setting of MS-specific targets for *S.* Enteritidis and *S.* Typhimurium occurrence in laying hens (Hald et al., 2012). Thus, the difference between the two models follows the logic that if one or more sources have their contribution to the overall burden reduced, other sources will automatically contribute relatively more, and the fact the pigs

were the second most important source in the EU and the main in eight MSs make it the most obvious "substitute".

The relative contribution of different food sources is expected to vary between countries, influenced by food production systems, food consumption and preparation habits, food trade and the epidemiology of the pathogen in different regions (Pires et al., 2011a). The comparison of country-specific results of the EU model to results of single-country models applied in the EU is possible only for a few countries. Microbial subtyping models have been applied to data from Sweden (Whalström et al., 2011), the United Kingdom and the Netherlands (Pires, 2009), but these included only domestic sources as categories (in the Swedish study, there is also an aggregated category of "imported foods"). The proportions attributed to Swedish layers, pigs and broilers were different in the two models; this is likely due to differences in the data used, level of subtyping and number of categories included. The Swedish model includes cattle, geese and wildlife, and also uses phage typing data, allowing for a better differentiation of cases of the most common serovars between sources. In both models, the number of positive isolates in all sources was very small, and thus results were particularly sensitive to the discriminatory power of the data used. On the other hand, the estimated main sources of cases were similar in both models: imported foods (6.4% Whalström / 7.7% EU model) and travel (82% Whalström / 76% EU, showing an agreement between models in the sources for which a larger amount of data is available. Results of the model developed using Dutch data (Pires, 2009) are generally in agreement with our results, both in the order of importance of the animal sources and in the values estimated. The largest differences were observed in a lower proportion of cases attributed to layers in the EU model, which could be explained by the fact that the Dutch study used data from 2006, before the already mentioned activities for harmonized control of *S*. Enteritidis and *S*. Typhimurium in the EU (Hald et al., 2012). Results of the UK model developed in Pires (2009) do not corroborate our results, showing broilers as the main source, followed by layers, pigs and turkeys. One possible reason for that is that around 54% of UK cases in the EU model had travel information recorded as "unknown", and were thus regarded as domestic, while at the same time, 42% of properly reported travelers were S. Enteritidis cases. So, if we are to believe that the predominance of this serovar is maintained among those that should have been recorded as travel-related, it is likely that the EU model has wrongly attributed travel cases to layers. This is further supported by the fact that the *Salmonella* control in eggs and layers in the UK started in 1989, with the establishing of the current program in 2004 (DEFRA, 2007), and the reported data for layers in UK have shown a low prevalence in many years (also before 2004, when the BS was conducted). In the data used for the EU model, 1.2% of monitoring samples were positive for *Salmonella*. We therefore believe that the results of the single-country model are more correct for this reservoir, as well as for travel-related cases. Comparisons with the Danish source account model have been extensively presented in section 6.4 and will be discussed in separate.

A large proportion of cases reported in Finland, Sweden, Ireland, the UK and Denmark were attributed to international travel, whereas travelling appeared to be less important in remaining countries, varying from 14% in the Netherlands to 0% in e.g. Spain. The Danish source account estimated a higher proportion of travel-related *Salmonella* cases, varying between 22 and 46% (Anonymous, 2008; Anonymous, 2009; Anonymous, 2010), but these have been estimated accounting for the probability of a case with unknown travel information having been travelling abroad before onset of symptoms, and thus added more "possible" travelers. In the EU model, it was not possible estimate additional "extra" travelers because the proportion of reported cases with missing travel information varied substantially, being 100% in some countries. Moreover, travel information as reported to TESSy is often incomplete and may not reflect the true relation between travel and domestic cases (EFSA, 2012a). Travel-related disease therefore corresponded solely to

the cases that were reported as acquired outside the country, assuming that all reported cases with missing travel information were domestically acquired, and is in general expected to be underestimated. As examples, Belgium, France, Spain, Romenia and Slovenia systematically report zero travel cases every year (EFSA, 2009a; EFSA, 2010a; EFSA, 2011a), even though there is literature describing travel-related cases of salmonellosis in both France (Weill et al., 2006; Hendriksen et al., 2012) and Belgium (Bottieu etl al., 2011). Finally, in MSs with reasonably good travel data it can be seen that a large proportion of the *S.* Enteritidis infections are linked to travel, which is an indication that the same could happen in the MSs which have poor or no travel data. This suggests that unreported travel-related cases would be wrongly attributed to one of the animal food sources included in the model, as also observed by Hald et al. (2012). It was not possible to differentiate between travelling within or outside Europe, since this information was only available for a few MSs.

In general, it is not possible to directly compare the proportion of cases attributed to outbreaks with the results of previously published source attribution studies. When looking at the UK, for example, the data used in the EU model had no outbreak cases reported (Table 5), while in Pires (2009), 5.8% of cases were allocated to this source. In the Netherlands and Sweden, on the other hand, the differences observed reflect the use of the UFs as multipliers for sporadic cases. As described in the methods section, it was assumed that outbreaks have a higher probability of being reported than sporadic cases, and so the same UFs could not be applied. It is therefore expected that in the EU model these cases represent a different proportion of the total burden when compared to other studies in which there was no correction for underreporting of sporadic cases. As an example, in the Netherlands, 10.8% of cases were attributed to outbreaks by Pires (2009), compared to 0.5% by the EU model; the UF for the Netherlands is 26.3, and if proportions are recalculated not taking that into consideration, Dutch outbreak cases in the EU model correspond to 11.3% of the total. In Sweden, where Whalström et al. (2011) reported 2.9% of cases as associated to outbreaks against 4.4% in the EU model, our proportions can again be recalculated taking into consideration the Swedish UF of 0.5, resulting in a non-adjusted attributed proportion of 2.3%. Therefore, although the models discussed are mostly focused on the attribution of sporadic cases, the proper reporting of outbreak cases is regarded as an important requirement for the accuracy of their results.

The attribution of human cases to a limited number of food-animal sources may result in the misplacing of some cases if their "true" source is not included. As an example, according to the GEMS/Food Cluster Diets (GEMS/Food, 2006), beef is the second or third most consumed food-animal in most EU MSs, with consumption being normally lower than pork, but varying in relation to broiler meat, depending on the country. Although *Salmonella* prevalences in beef and beef products are normally low in the EU (EFSA, 2012a) the non-inclusion of the cattle reservoir in the EU model is likely to have resulted in some beef-related cases being "wrongly" attributed to pigs, as *S.* Typhimurium is one of the main serovars in both sources. Nonetheless, the initial Danish model only included five sources, and it was still a powerful tool in guiding the decisions for the targeted actions regarding broilers, pigs and table eggs that dramatically decreased the prevalence of *Salmonella* in these sources in the last decade (Anonymous, 1998; Wegener et al., 2003).

Other foods recognized as sources of human salmonellosis, such as fruits and vegetables, were not included in the model. However, it should be highlighted that the subtyping approach employed is tracing human infections back to the animal reservoir. This means that human infections caused by fruits and vegetables contaminated with feces from food-producing animals would be traced back to this reservoir, which may be useful for some types of risk management decisions. Still, there is evidence that *Salmonella*-

contaminated foodstuffs are imported into EU from third countries. Such foodstuffs obviously constitute a risk for humans, but their relative importance could not be accounted for in the model.

8.1.1. Data management and selection

Data used in the source attribution model were retrieved from multiple sources and presented varied levels of quality and completeness. Although TESSy and EFSA collect and organize the data at EU-level in a harmonized way, the primary information is collected in different countries, which have their particular approaches and methods for data collection and management. This variability would affect the model's design and output and made several data management steps necessary.These steps have by their turn an impact on the final estimates and are here discussed.

Differences include different levels of underreporting, which were taken into consideration by the use of underreporting factors. Limitations and assumptions of the use of those factors should be discussed, as they were calculated based on Swedish cases (de Jong and Ekdahl, 2006; Havelaat et al., 2012), which came from a system where underreporting is also expected to occur. Also, by using the infection rates in returning travelers to calculate incidences for the local population in the countries visited, it was assumed that the eating habits and other exposures of Swedish travelers are the same as the locals', also disregarding local levels of acquired immunity and differences in circulating strains. Similar considerations must be done regarding the use of a Dutch population-based *Salmonella* prevalence study as a reference to estimate the underreporting in the other countries. A full discussion of the limitations, as well as a comparison of these estimates with the ones from 2006 can be found in Havelaar et al. (2012). Although the 2012 UF values are different from the ones from 2006 (Havelaar et al., 2012), there is a high correlation between the probability of cases being reported on arrival in the two studies and also between the incidence rates found in both, so the UFs based on the most recent data were considered validated enough to be applied to the raw numbers reported to TESSy.

The use of underreporting factors has proved important when considering the effect of source and country contributions at EU level. This is particularly clear for broilers: this reservoir was the most important only in Portugal, but the use of an underreporting factor multiplied its impact within the EU by 2082.9, increasing both the relative contribution of broilers and of Portugal to the total cases of salmonellosis in the EU, when compared to the original numbers. A similar effect can be observed for the contribution of Greece to the total cases attributed to layers. However, it is noticeable that most of the cases "originated" by countries with large underreporting factors were reported in those same countries, so one should be careful when interpreting these results as countries "exporting" cases to the rest of the EU.

Given the differences in the frequency of reporting among countries, it was necessary to sum human data from the years of 2007, 2008 and 2009, in order to obtain a more robust dataset to work with. This was also done because animal BS data were collected in different years, so in order to obtain a dataset with a temporal relation between animal prevalences and human cases this data aggregation step was assessed to be necessary. Results therefore do not apply to individual years, and this single model does not have the objective of observing trends over time. Variability of outbreaks during years also do not affect the model results, as outbreaks are removed from the total cases, summed in separate and presented as a category labeled "outbreaks", not having any influence on the attribution to sources.

Concerning the animal data, the panel of participating MSs varied with each BS, as countries have the right to refuse participation in the EU-wide Baseline Studies. The admittance of new MSs to the EU also generates different lists of reporting countries for each animal source, as data were collected in different

years. The resulting data gaps were, when possible, filled with information from the EFSA EUSR. There are currently no EU-wide studies on the baseline prevalences of *Salmonella* in cattle and no harmonized monitoring in place, which is the main reason why this reservoir was excluded from this study.

Data were also heterogeneous in regards to serotyping information and reporting of aggregated data or data with no or sparse serotyping information for both humans and animals. To deal with missing or aggregated information, records were reassigned based on specific criteria, and countries were approached directly for more complete datasets. Reassignment was based on the serovar distributions observed in available data or external reference datasets (e.g., WHO GFN/CDB), and this approach obviously has limitations. Any emergence of new serovars or other profile fluctuations may be lost, particularly in situations where a whole year of typing is missing and the records are reassigned based on data from previous years. Therefore, the serovar reassignment is considered a large source of uncertainty around the final data, and the model could benefit from a reassigning approach that uses a stochastic process, allowing for this uncertainty to be expressed and quantified. At the current stage, due to the amount of different scenarios of non-identification of serovars (Figure 3) and the need to use data from different external sources, developing such an approach was not possible. This should, however, be attempted in future versions of the model, particularly if establishing a EU-harmonized surveillance (ECDC, 2007; EC Decision 2002/253/EC, 2008) improves the homogeneity of the data.

The prevalence data retrieved from the BS was included in the model as point estimates, and this represented another possible source of uncertainty. Two possibilities to include the uncertainty around these data in the model were investigated. The first was to have the prevalences simulated from prior distributions defined based on the point estimates and the confidence intervals. This approach would allow the use of the weighted prevalences available in the study reports and their calculated uncertainty. The approach was not applicable because weighted prevalences were only calculated for a number of serovars, and so confidence intervals were not available for all serovars included in the model. The second approach was to follow the methodology described by Müllner et al. (2009), in which the prevalence data was fit as a beta distribution with the mean and the standard deviation of the prevalence as parameters α and β. These parameters were calculated based on simulations of the overall *Salmonella* prevalence and the relative occurrence of serovars, and the approach was justifiable because the data used by Müllner et al. derived from multiple data sources and were not representative of the study population. However, this approach leads to an overparametrization of the model because the model has only two data points (the prevalence and the amount of a food source available for consumption), and was therefore not used. Still, the BS data were considered reliable because the target sample in all baseline studies was 80% of the total investigated epidemiological unit (herd, holding, flock or unit, depending on the species), and a strict process to replace missing sample points and to exclude non-conformant collected samples was performed to achieve a representative sample at both country and EU-level in all studies (EFSA, 2007b; EFSA, 2008a; EFSA, 2008b; EFSA, 2010c). The same apply to the laying hens data, as the EU-harmonized monitoring includes all commercial flocks of laying hens, broilers and turkeys.

The lack of further subtyping information on *S.* Enteritidis and *S.* Typhimurium is likely to have resulted in attribution of some human cases to the wrong source. In MSs where *S.* Enteritidis is prevalent in both slaughter pigs and laying hens, the current subtyping level makes it difficult for the model to distinguish between the two sources. Some *S.* Enteritidis cases may therefore have been wrongly attributed to pigs instead of laying hens, and it is also likely that the number of broiler-related *S.* Enteritidis infections has been overestimated for the same reason, which was also observed in Hald et al. (2012). However, given the current availability of data, it was considered that the use of countries as a third dimension and differences in

serovar distribution in the sources among countries can to some extent compensate the low discriminatory power of using only serotyping Pires and Hald, 2010).

Data with higher discrimatory power (e.g. phage typing or antimicrobial resistance susceptibility testing) were unavailable because these methods are not routinely applied in several of the included countries. Because it is not likely that phage typing or antimicrobial resistance profiles will become available at a general level in the EU, the use of other typing methodologies should be explored.One future possibility for improving the level of subtyping detail is by applying genotyping methods, which use is expected to increase in the next years (Chenu et al., 2012). Those methods produce fast results and are becoming increasingly cheaper, which should allow countries to subtype and submit a larger amount of properly identified isolates. DNA-based molecular methods have a higher discriminatory power than phenotypic methods (Chenu et al., 2012), which makes them particularly useful for outbreak investigations, when there is a need to pinpoint a particular source (Torpdahl et al., 2007; Baggesen et al., 2010). Their utility for source attribution of sporadic salmonellosis is still to be explored. The use of genotypic subtyping methods would theoretically allow the adaptation of the Asymmetric Island model for *Salmonella*, as was done for *Campylobacter* in England (Wilson et al., 2008), New Zealand (Müllner et al., 2009) and Denmark (Boysen, 2012). This approach treats animal and environmental sources of the pathogen as populations among which there may be gene migration. In each of those populations, the bacteria evolve independently through new mutations or horizontal gene transfer (recombination). Mutation, migration and recombination rates are then estimated and used to assign human cases probabilistically to one of the source populations (Wilson et al., 2008). The applicability of this approach for *Salmonella* has thus far not been tested because appropriate genotyping data (e.g. Multiple-Locus Variable number tandem repeat Analysis - MLVA) data are not yet available. The Island Model has the advantage of not requiring a full match between human and source isolates, thus making it possible to attribute human isolates not observed in the reservoirs to the most likely source. However, the probability-based attribution means that cases will be directed to the source in which the highest probability of origin was found, even if that probability is low (Boysen, 2012.), not allowing the existence of a "unknown source" category. So, although those methods are a promising tool, there still needs to be some evaluation of the appropriate level of discrimination that is useful for source attribution based on microbial subtyping.

The attribution estimates took into account the amount of food produced and traded between countries as reported to the EUROSTAT database. The underlying assumptions were that the EUROSTAT data were complete and consistent, that not all food produced in a country is exported and that all the food available for consumption is actually consumed, in a way that these data reflected the real flow of foodstuffs and consequent exposure in the countries. These were strong assumptions, as also stated by Vose et al. (2011) and Hald et al. (2012), and these data presented a special challenge, as they had to be built based on four primary EUROSTAT datasets. According to a quality assessment performed by EFSA (2010f), the information recorded in those datasets does not fully support our assumptions. This assessment showed the existence and non-reporting of triangular trade, mis-classification of food products and problems in the conversion of currency/weight units. Also, in several situations, data had to be estimated for missing years or supplied with further surrogate data (e.g. AVEC data). It is an important feature in this model that the relative contribution of food-animals produced in different countries is dependent not only on the *Salmonella* prevalence in a source in an exporting country, but also on the amount imported from that country. This is a point in which the EU model differs from the way single-country models work: in a single-country model, m_j works as a subset of a_j, as they have the same dimensions (Hald, 2004; Pires and Hald, 2010; Whälstrom, 2011); for each source, there is only one value of m_j and one value for the prevalence (p_{ji}) of a subtype in that

source. As a consequence, $m_{j,}$ has the role of weighting the contribution of the different sources, which is to some extent also reflected in a_j. In the multi-country model, *m* in a reporting country is composed by subsets of *m* from different countries of origin of the food sources, each one with its own prevalence. For that reason, even if an exporting country has a very high prevalence in a source, this prevalence will have little impact in an importing country if the amount imported is very small, particularly if another country with a low prevalence exports very large amounts which can "dilute" the high prevalence found in the first country. In short, the amount imported ultimately drives the *m*p* in the model formula, particularly when large differences in trade volume are observed. In order to assure that the use of these data would not compromise the model results, data were compared with the WHO GEMS/Food data for validation. This assessment revealed that the chosen data management produced data which were in line with the consumption data in the GEMS dataset.

8.1.2. Models comparison

The comparison between Danish source attribution estimates obtained by the EU *Salmonella* source attribution model and by the single-country model applied using Danish data only was performed to assessed the impact of differences between the two models and conclude on advantages and limitations of each. Differences derive mostly from the type of data used, and reflect the different levels of subtyping, as well as the inclusion of different sources. These two differences have different impacts on the results.

As an example, imported foods are not explicitely included as a category in the EU model, but are present as part of each source category, as the imported volumes and the prevalences in the exporting countries are taken into consideration. The impact of this difference is particularly evident for eggs of foreign origin, which are not monitored in Denmark and therefore not included in the Danish model. In Denmark, human cases are assumed to originate from domestic eggs only because the national food authorities and the industry consider that the vast majority of imported eggs are only used for processed foods that undergo heat-treatment. This type of country-specific information was not available for the EU study, but future EU models could include such data and adjust the results accordingly; in the above example, this would imply disregarding the cases attributed to imported eggs.

When compared to the EU model, the use of the case-by-case data in the Danish model results in an underestimation of the contribution of imported meats to the *Salmonella* cases reported in Denmark because it does not sample foods from all countries exporting to Denmark. However, data from the main contributing countries in all categories (particularly for pigs, where most cases are domestic) are available, as well as data from products imported from non-EU countries, suggesting that the case-by-case data has a good level of sensitivity and representativeness for the purposes of the DK model.

The models's results also differed in the amount of cases attributed to an "unknown" source, which reflects the different number of sources in the two models and the higher level of subtyping detail in the DK model. Because the EU model attributes to a lower number of sources and uses data with lower discriminatory power, the proportion of cases with an unknown source is expected to be higher. Although existent, this difference was not substantial, and this is probably due to the attribution of cases to one of the included sources where frequent serovars were also isolated.

The proportion of cases attributed to outbreaks differed substantially in the two models because the attribution estimates for all sources except outbreaks in the EU model were adjusted for underreporting. This changed the balance between the proportion of cases attributed to outbreaks and to the other categories, when

comparing the two models. Re-calculation of the proportions after readjustment of the estimates of the two models would reveal similar estimates for this category.

Results of the single-country model could be improved by the use of country-specific trade data for the *m* component, thus taking into consideration a weighted contribution of exporting countries to the number of cases attributed to the sources. Two limitations are readily visible in this scenario, as specific imported amounts would require country-specific positive percentages or prevalences for the model: 1) if we are to trust the trade data used for the EU model, although the main contributing countries are represented, results from the case-by-case study are still not fully representative of the variety of countries from which Denmark imports food; 2) data from the EFSA baseline studies as well as newer data from the EU harmonizing monitoring programs lack the required subtyping detail to keep the level of discriminatory power that the model currently has.Thus, here is currently no "better" data to replace the case-by-case.

Despite the discussed data limitations and differences, results of the EU model seem to point in the same direction as the Danish model for prioritizing interventions at the national farm-to-fork chain, showing almost the same order of importance for the sources common to both. The main difference was observed for turkeys, and it was not possible to evaluate which of the models present a more realistic estimation.

8.2. Source attribution using expert elicitation based on cluster analysis

In this pilot, estimates for attributable proportions of *Salmonella* cases to broilers, pigs, layers, turkeys and travel in Romania, Bulgaria, Norway and the Czech Republic were obtained through expert elicitation. This approach was attempted because the required data to include those countries in the microbial subtyping-based model were not available. Attribution results and serovar occurrence in the four sources were used as basis for clustering countries present in the EU model, characterizing "profile groups"; then, sets of non-health-related demographic, economic, climate and animal production data were used to cluster those same countries with Romania, Bulgaria and Norway, creating non-health-based "profile groups". A panel of experts was then asked to analyze how countries fitted in the *Salmonella* profile groups and how they fitted together with the added countries in the non-health profiles, and use that information as basis for their estimates.

When considering cases attributed to sources, the majority of *Salmonella* cases for the Czech Republic were attributed to layers, followed by pigs, turkeys / travel and broilers. This order of importance is well in line with the estimates from the EU model, although the values were different. The largest difference was observed for the "unknown" category. However, as presented in the evaluation of the method by the experts, this category received the cases which were not destined to any other categories, becoming a "depository of uncertainty" instead of a true attribution category.

The validation of this of approach was difficult because there are no other source attribution estimates to compare with the estimates for Bulgaria, Norway and Romania, and the ones used for the Czech Republic were also obtained from a model that is still being evaluated and under a process of improvement. Even so, a proportionality similarity index calculated for the two sets of Czech results showed a high level of similarity between them, particularly if the "unknown" category is removed because of the aforementioned reasons. As for the remaining three countries, a visual assessment of the behavior of the panel concerning homogeneity and uncertainty of guesses was conducted, and they seemed to generally follow the same pattern observed for the Czech Republic.

Although the piloted approach still needs improvements, results suggested that this novel method could be useful to assess the relative importance of food sources and define priorities for intervention. This pilot study was useful to identify limitations of the current method and potential solutions or alternative approaches for modifications.

One of the basic assumptions of the method was that socioeconomic and climate data could be used as input for a source attribution analysis in countries with poor *Salmonella* data. Nonetheless, this assumption was not supported by the experts, who did not not use these data at any moment. This may have been due to the fact that the members of the panel already worked in EU-wide projects, knowing the countries under study reasonably well, thus not requiring this type of background information. To assess if these data could be useful in other studies, future attempts should include a more international panel, whose members could potentially find the socio-economical profiling more necessary. In any case, the type of expertise the panel showed is an added value for this type of elicitation.

An alternative way to evaluate this approach would be to adapt the methodology and materials to use the attribution results for Latin America and the Caribbean obtained by Pires et al. (2012) and try to extrapolate them to surrounding countries not included in that model. If a different panel in a different area reaches results as agreeable as in this pilot, this could mean source attribution based on expert elicitation informed by cluster analysis of non-health variables is a useful approach for obtaining source attribution estimates on a more global scale.

A more conservative approach to attribute foodborne illnesses in countries where no sufficient data are available for a source attribution approach would be to adapt the regression model used by Hansen (2012). This study investigated the usefulness of non-health predictors to predict foodborne disease-related mortality in multiple countries. In our proposal, the purpose would be to estimate the correlation between each predictor and the proportion of disease attributed to food sources, and then apply this effect to the predictor values observed for the countries without attribution results. That would require adapting a multinomial regression model with probabilities as outcomes, since we intend to predict the different attributable parcels simultaneously. Alternatively, one linear regression model could be run per source. At this point, neither of these approaches were applied because we had only 24 observations, and the sample size was considered too small to obtain any valid results. This approach can be attempted when traditional attribution estimates from a larger number of countries are available.

8.3. General discussion

The application of a microbial subtyping approach for source attribution of human salmonellosis at the EU level was successful and produced novel and useful results. We considered that the management approach applied to the available data produced datasets considered useful for the applied source attribution method, provided that a thorough data evaluation of the data is performed and specific countries and reservoirs with insufficiently representative data are excluded, thus accomplishing objective 4.1.1. This shows that, as long as existing limitations are taken into consideration and clearly reported, public surveillance and monitoring data can be used for scientific purposes, and this could be a first step to the conduction of source attribution studies in countries where no country-wide baseline studies or serovar surveys have been conducted, but where programs for *Salmonella* monitoring in food or surveillance in humans are currently running.

Despite data limitations and the consequent uncertainty in the results, the source attribution estimates are considered valid as a first indication of which sources are most important for human salmonellosis in several

countries. Limitations include the variability in the human surveillance systems in place in the countries, as well as the different details with which serovar information is reported for both human and animal-food sources. Such uncertainties cannot be statistically quantified, but should be kept in mind when interpreting the results. The relative importance of different food-animal sources was found to vary between countries according to differences in prevalences, trade and consumption patterns and preferences, as well as animal and food production systems, also highlighting regional differences in the focus of surveillance systems in place in EU Member States. Thus, objective 4.1.2. was considered as accomplished, as the results of the model are expected to be useful for the delineation of risk management strategies in the EU, particularly if it is applied on a regular basis, to evaluate the impact of targeted interventions and dynamic changes in the sources of human salmonellosis. As a consequence of the accomplishing of objective 4.1.3., improvements to the model have been proposed and implemented. A good example of that is shown in Hald et al. (2012), where the decrease of egg-related cases in the EU due to control measures can be observed from the application of updated data to the EU model. Also, a user-friendly tool for running the EU-model, as well as country-specific source attribution, was recently developed for EFSA, making this procedure more accessible for both EFSA and a larger number of countries (Hald and Lund, 2012).

Concerning the proposal of an alternative approach for source attribution for countries with missing data, we considered that objective 4.2.1. was accomplished, as the potential usefulness and viability of the clusters-based expert elicitation are visible. However, the right combination of non-health variables is still to be found. As of now, the main difficulty in the application of this approach in a global scale is that available source attribution estimates for different countries must have been obtained under the same study, or at least focused on the same sources; otherwise, the clustering process gets compromised. Notwithstanding, we suggest that modifications to the applied methodology could improve the approach and achieve better results.

9. CONCLUSION

We conclude that Hypothesis 1 (It is possible to develop an EU model based on the data available) is accepted, as the development of a useful EU model with the available EU surveillance data was successful, and this proved to be a viable option for countries with less intensive data-collection systems than established e.g. in Denmark.

On the contrary, Hypothesis 2 (It is possible to extrapolate results of the EU model to countries with insufficient data using non-health indicators and expert elicitation), is at this point rejected, as the non-health indicators *per se* were not used by the panel of experts.

10. PERSPECTIVES

Several improvements to the two approaches could be applied. Proposed improvements to the EU microbial subtyping model include:

a) Add a stochastic feature to the serovar reassigning, which would allow for the expression of the uncertainty inherent to the process.
b) Develop an approach to use animal prevalence data as a stochastic node, in an attempt to allow for the model to estimate prevalences for countries where those data is not available.
c) If the model is not able to estimate the prevalences for lack of input data, use the same materials from the clusters EE to elicit a) surrogate prevalence data for the countries and sources which are missing and include them in the EU model, or b) surrogate values for *a* and *q*, and have the model estimate the prevalences.

Additionally, we expect that the potential for repeating the application of the EU model in upcoming years should motivate countries to improve the reporting of isolates. Although it is only mandatory for MSs to report *S.* Enteritidis and *S.* Typhimurium, subtyping of all isolates is performed to identify those two, and so other serovars can be easily reported.

Another interesting prospect is to set up a research initiative looking into genotypic sequencing methods (MLVA, MLST, other) which are more suitable for source attribution, including an assessment of the discriminatory level that makes the most appropriate distinction between epidemiologically related and non-related strains. On a more long-term basis, that could result in the setting up of a system collecting genotypic information from the harmonized EU *Salmonella* surveillance programmes, so that future source attribution studies could be based on genotypic subtypes. Finally, further application and evaluation of the subtyping approach for other foodborne pathogens should be pursued in future studies.

As for the clusters elicitation, perspectives include the conduction of a full study after this pilot, using a larger panel, and also the application of the same model to Latin America and the Caribbean. The prospect is to use an international panel and posteriorly comparing the uncertainty level provided by experts in the two models.

11. REFERENCES

Adams, M.R., Moss, M.O. 1995. Factors affecting the growth and survival of microorganisms in food. In: Food microbiology. Cambridge: The Royal Society of Chemistry, 42-43. Il.

Acha, P.N.; Szyfres, B. 2011. Zoonosis y enfermedades transmisibles comunes al hombre y a los animales. 3ª ed. Washington: Organización Panamericana de la Salud. p. 233-238. (Publ. Cient. nº 580)

Anonymous. 1998. Annual report on zoonoses in Denmark 1997. Copenhagen, Denmark: The Ministry of Food, Agriculture and Fisheries. Available from: http://www.vetinst.dk.

Anonymous. 1999. Annual report on zoonoses in Denmark 1998. Copenhagen, Denmark: The Ministry of Food, Agriculture and Fisheries. Available from: http://www.vetinst.dk.

Anonymous. 2000. Annual report on zoonoses in Denmark 1999. Copenhagen, Denmark: The Ministry of Food, Agriculture and Fisheries. Available from: http://www.vetinst.dk.

Anonymous. 2012. Annual report on zoonoses in Denmark 2011. National Food Institute, Technical University of Denmark. ISSN 0909-3837.

AVEC, 2009. Annual Report 2009. Association of Poultry Processors and Poultry Trade in the EU Countries, Brussels, Belgium

Baggesen, D.L., Sørensen, G., Nielsen, E.M., Wegener, H.C. 2010. Phage typing of *Salmonella* Typhimurium – is it still a useful tool for surveillance and outbreak investigation? Euro surveillance 15(4),19471.

Baker, M.G., Thornley, C.N., Lopez, L.D., Garrett, N.K., Nicol, C.M., 2007. A recurring salmonellosis epidemic in New Zealand linked to contact with sheep. Epidemiol. Infect. 135, 76-83.

Batz, M.B., Doyle, M.P., Morris, G.M., Painter, J., Singh, R., Tauxe, R.V., Taylor, M.R., Lo Fo Wong, D.M.A. 2005. Attributing illness to food. Emerging Infectious Diseases 11(7), 993-999.

Bottieau, E., et al. 2011. Epidemiology and outcome of Shigella, Salmonella and Campylobacter infections in travelers returning from the tropics with fever and diarrhoea. Acta Clinica Belgica 66(3):191-195.

Boysen, L. 2012. *Campylobacter* in Denmark – Control, human risk and source attribution. PhD Thesis. Technical University of Denmark. ISBN: 978-87-92763-22-8.

Buettneri, S., Wieland, B., Staerk, K.D.C., Regula, G. 2010. Risk attribution of *Campylobacter* infection by age group using exposure modelling. Epidemiol. Infect. 138, 1748–1761.

Chenu, J.W., Cox, J.M., Pavic, A. 2012. Classification of *Salmonella* enterica serotypes from Australian poultry using repetitive sequence-based PCR. Journal of Applied Microbiology 112(1), 185-196.

Cooke, R.M., Goossens, L.H.J. 2004. Expert judgment elicitation for risk assessments of critical infrastructures. Journal of Risk Research 7(6), 643-656.

David, J. 2009. Attribution des cas de salmonelloses humaines aux différentes filières de production animale en France: Adaptabilité et robustesse du modèle bayésien d'attribution par typage microbiologique. PhD Thesis. Université Européenne de Bretagne. N° Ordre : 2009-25.N° Série : B-201.

Davidson V.J., Ravel A., Nguyen T.N., Fazil A., Ruzante J. 2011. Food-specific attribution of selected gastrointestinal illnesses: estimates from a Canadian expert elicitation survey. Foodborne Pathogens and Disease. 8(9), 983-995.

De Irala-Estevez, J., Groth, M., Johansson, L., Oltersdorf, U., Prättälä, R., Martinez-González, M.A. 2000. A systematic review of socio-economic differences in food habits in Europe: consumption of fruits and vegetables. European Journal of Clinical Nutrition 54, 706-714.

De Jong, B., Ekdahl, K. 2006. The comparative burden of salmonellosis in the European Union member states, associated and candidate countries. BMC Public Health 6,4.

DEFRA. 2007. UK National Control Programme for Salmonella in layers (*Gallus gallus*). London, United Kigdom: Department for Environment, Food and Rural Affairs. 30p.

Domingues, A.R., Pires, S.M., Halasa, R., Hald, T. 2012. Source attribution of human salmonellosis using a meta-analysis of case-control studies of sporadic infections. Epidemiol. Infect. 140, 959–969.

ECDC. The European Centre for Disease Prevention and Control. 2007. Surveillance of communicable diseases in the European Union: A long-term strategy 2008–2013. Acessed on November 2012. Available at http://ecdc.europa.eu/en/aboutus/Key%20Documents/08-13_KD_Surveillance_of_CD.pdf.

EFSA. 2007a. The Community Summary Report on Trends and Sources of Zoonoses, Zoonotic Agents and Food-borne Outbreaks in 2006. EFSA Journal 130(3):352.

EFSA. 2007b. Report of the Task Force on Zoonoses Data Collection on the Analysis of the baseline survey on the prevalence of *Salmonella* in broiler flocks of *Gallus gallus*, in the EU, 2005-2006. Part A: *Salmonella* prevalence estimates. The EFSA Journal 98, 1-85.

EFSA. 2007c. Report on the Analysis of the baseline study on the prevalence of *Salmonella* in holdings of laying hen flocks of *Gallus gallus*. EFSA Journal 97.

EFSA. 2008a. Report of the Task Force on Zoonoses Data Collection on the analysis of the baseline survey on the prevalence of *Salmonella* in slaughter pigs, in the EU, 2006-2007. Part A: *Salmonella* prevalence estimates. Question N° EFSA-Q-2006-042A. The EFSA Journal 135.

EFSA. 2008b. Report of the Task Force on Zoonoses Data Collection on the Analysis of the baseline survey on the prevalence of *Salmonella* in turkey flocks, in the EU, 2006-2007. Part A: *Salmonella* prevalence estimates. The EFSA Journal 134.

EFSA. 2008c. Scientific opinion of the panel on Biological Hazards on a request from EFSA on Overview of methods for source attribution for human illness from food borne microbiological hazards. The EFSA Journal 764, 1-43.

EFSA. 2009a. The Community Summary Report on Trends and Sources of Zoonoses, Zoonotic Agents and Food-borne Outbreaks in 2007. EFSA Journal 223(7):312.

EFSA. 2009b. Scientific Opinion of the Panel on Biological Hazards on a request from European Commission on Quantitative estimation of the impact of setting a new target for the reduction of *Salmonella* in breeding hens of *Gallus gallus*. The EFSA Journal 1036:1-68.

EFSA. 2010a. The Community Summary Report on Trends and Sources of Zoonoses, Zoonotic Agents and Food-borne Outbreaks in the European Union in 2008, EFSA Journal 8(1):1496

EFSA. 2010b. Panel on Biological Hazards (BIOHAZ); Scientific Opinion on monitoring and assessment of the public health risk of "*Salmonella* Typhimurium-like" strains. EFSA Journal 2010;8(10):1826.

EFSA. 2010c. Analysis of the baseline survey on the prevalence of Campylobacter in broiler batches and of Campylobacter and *Salmonella* on broiler carcasses in the EU, 2008 - Part A: Campylobacter and *Salmonella* prevalence estimates. EFSA Journal 8(3):1503.

EFSA. 2010d. Manual for Reporting on Zoonoses, Zoonotic Agents and Antimicrobial Resistance in the framework of Directive 2003/99/EC and of some other pathogenic microbiological agents for information derived from the reporting year 2009. European Food Safety Authority.

EFSA. 2010e. Panel of Biological Hazards (BIOHAZ); Scientific Opinion on a quantitative risk assessment of Salmonella in slaughter and breeder pigs. EFSA Journal 8(4):1547.

EFSA. 2010f. Collection and routine analysis of import surveillance data with a view to identification of emerging risks. EFSA Journal 8(3):1531.

EFSA. 2011a. The European Union Summary Report on Trends and Sources of Zoonoses, Zoonotic Agents and Food-borne Outbreaks in 2009. EFSA Journal 9(3):2090.

EFSA. 2011b. EFSA Panel of Biological Hazards (BIOHAZ); Scientific Opinion on a quantitative estimation of the public health impact of setting a new target for the reduction of *Salmonella* in broilers. EFSA Journal 9(7):2106.

EFSA. 2012a. The European Union Summary Report on Trends and Sources of Zoonoses, Zoonotic Agents and Food-borne Outbreaks in 2010. EFSA Journal 10(3):2597.

EFSA. 2012b. EFSA Panel of Biological Hazards (BIOHAZ); Scientific Opinion on an estimation of the public health impact of setting a new target for the reduction of *Salmonella* in turkeys. EFSA Journal 10(4):2616.

Ethelberg, S., Wingstrand, A., Jensen, T., Sørensen, G., Müller, L., Nielsen, E.M. 2008. Large ongoing outbreak of infection with Salmonella Typhimurium U292 in Denmark, February-July 2008. Euro Surveillance 13(28):pii= 18923. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=18923.

EUROSTAT. The Statistical Office of the European Union. Accessed on June 15, 2011. Available at http://epp.eurostat.ec.europa.eu/portal/page/portal/about_eurostat/corporate/introduction

Evers, E.G., Van Der Fels-Klerx, H. J., Nauta, M.J., Schijven, J.F. 2008. Campylobacter source attribution by exposure assessment. International Journal of Risk Assessment and Management 8(1), 174.

Feinsinger, P., E. E. Spears, and R. W. Poole. 1981. A simple measure of niche breadth. Ecology 62:27–32.

Food and Agriculture Organization of the United Nations /World Organization for Animal Health. 2009. Guide to Good Farming Practices. ISBN 978-92-5-006145-0. 22p.

Food and Agriculture Organization of the United Nations. 1995. Feed commodity demand and supply. In: Interactions between Livestock Production Systems and the Environment - Impact Domain: concentrate feed demand. FAO Corporate Document Repository. Available at http://www.fao.org/wairdocs/lead/x6123e/x6123e00.htm. Accessed on November 15, 2011.

Gebreyes, W. A., Altier, C., Thakur, S. 2006. Molecular epidemiology and diversity of *Salmonella* serovar Typhimurium in pigs using phenotypic and genotypic approaches. Epidemiol. Infect. 134, 187–198.

Gebreyes, W. A., Thakur, S. Phenotypic subtyping of foodborne pathogens. In Brul, S., Fratamico, P.M., McMeekin, T. A. (eds.). Tracing pathogens in the food chain. Woodhead Publishing Series in Food Science, Technology and Nutrition: Number 196. 141-156. 2011.

Gelman, A., Rubin, D.B. 1992. Inference from Iterative Simulation Using Multiple Sequences. Statistical Science 7(4), 457-472.

GEMS/Food regional diets: regional per capita consumption of raw and semi-processed agricultural commodities / prepared by the Global Environment Monitoring System/Food Contamination Monitoring and Assessment Programme (GEMS/Food). -- Rev. ed.

Gilks, W.R., Richardson, S., Spiegelhalter, D. (eds.). 1996. Markov Chain Monte Carlo in practice. New York:Chapman and Hall/CRC Interdisciplinary Statistics. 512p. Ill.

Greger, M. 2007. The human/animal interface: emergence and resurgence of zoonotic infectious diseases. Crit Rev. Microbiol. 33, 243-299.

Guo, C., Hoekstra, R. M., Schroeder, C. M., Pires, S.M., Ong, K.L., Hartnett, E., Naugle, A., Harman, J., Bennett, P., Cieslak, P., Scallan, E., Rose, B., Holt, K.G., Kissler, B., Mbandi, E., Roodsari, R., Angulo, F. J., Cole, D. 2011. Application of Bayesian techniques to model the burden of human salmonellosis attributable to U.S. food commodities at the point of processing: adaptation of a Danish model. Foodborne Pathog. Dis. 18(4), 509-516.

Haesebrouck, F., Heyndrickx, M., Pasmans, F., Ducatelle, R., Decostere, A. 2005. Recent changes in *Salmonella* nomenclature: The need for clarification. The Veterinary Journal 170, 275–277.

Hald, T., Vose, D., Wegener, H.C., Koupeev, T., 2004. A Bayesian approach to quantify the contribution of animal-food sources to human salmonellosis. Risk Analysis 24, 255-269.

Hald, T., Pires, S.M., de Knegt, L.V. 2012. Development of a *Salmonella* source-attribution model for evaluating targets in the turkey meat production. External Scientific Report Submitted to EFSA. EFSA-Q-2011-00341. Supporting Publications 2012:EN-259.35pp. Available online at: www.efsa.europa.eu/publications.

Hald, T., Lund, J. 2012. Development of a user-friendly interface version of the *Salmonella* source-attribution model. External Scientific Report Submitted to EFSA. EFSA-Q-2011-00341. *Awaiting publication.*

Hanson, L.A., Zahn, E.A., Wild, S.R., Döpfer, D., Scott, J., Stein, C. 2012. Estimating global mortality from potentially foodborne diseases: an analysis using vital registration data. Population Health Metrics 10(5), 1-7.

Havelaar, A., Bräunig, J., Christiansen, K., Cornu, M., Hald, T., Mangen, M-J.J., Mølbak, K., Pielaat, A., Snary, E., Van Pelt, W., Velthuis, A. Whalström, H. 2007. Towards an integrated approach in supporting microbiological food safety decisions. Zoonosis and Public Health 54, 103-107.

Havelaar, A.H., Galindo, A.V., Kurowicka, D., Cooke, R.M. 2008. Attribution of foodborne pathogens using structured expert elicitation. Foodborne Pathogens and Disease 5(5), 649-659.

Havelaar, A.H., Haagsma, J.A., Mangen, M.J., Kemmeren, J.J., Verhoef, L.P.B., Vijgen, S.M.C., Wilson, M., Friesema, I.H.M.,Kortbeek, L.M., van Duynhoven, Y.T.H.P., van Pelt, W. 2012a.

Disease burden of foodborne pathogens in the Netherlands, 2009. International Journal of Food Microbiology 156:231–238.

Havelaar, A.H., Ivarsson, S., Löfdahl, M., Nauta, M.J. 2012b. Estimating the true incidence of campylobacteriosis and salmonellosis in the European Union, 2009. Epidemiology and Infection 2012; 141: 293-302.

Hendriksen, R.S. 2010. Global epidemiology of non-typhoidal Salmonella infections in humans. PhD Thesis. Technical University of Denmark. ISBN: 978-87-92158-61-1.

Hendriksen, R.S., Vieira, A.R., Karlsmose, S., Lo Fo Wong, D.M.A., Jensen, A.B., Wegener, H.C., Aarestrup, F.M. 2011. Global Monitoring of *Salmonella* Serovar Distribution from the World Health Organization Global Foodborne Infections Network Country Data Bank: Results of Quality Assured Laboratories from 2001 to 2007. Foodborne Pathogens and Disease 8(8), 887-900.

Hendriksen, R.S., et al. 2012. Characterization of Isolates of Salmonella enterica Serovar Stanley, a Serovar Endemic to Asia and Associated with Travel. Journal of Clinical Microbiology 50(3):709-720.

Hoffmann, S., Fischbeck, P., Krupnick, A., McWilliams, M. 2007a. Using Expert Elicitation To Link Foodborne Illnesses in the United States to Foods. Journal of Food Protection 70(5), 1220–1229.

Hoffmann, S., Fischbeck, P., Krupnick, A., McWilliams, M. 2007b. Elicitation from large, heterogeneous expert panels: using multiple uncertainty measures to characterize information quality for decision analysis. Decision analysis 4(2), 91-109.

King, R., Brooks, S.P., Morgan, B.J.T., Gimenez, O. 2010. Bayesian Analysis for Population Ecology. Boca Ratón:Chapman and Hall/CRC 2009. 434p.

King, N., Lake, R., Campbell, D. 2011. Source Attribution of Nontyphoid Salmonellosis in New Zealand Using Outbreak Surveillance Data. Journal of Food Protection 74(3), 438-445.

Kingsley, R.A., Baumler, A.J. 2000. Host adaptation and the emergence of infectious disease: the *Salmonella* paradigm. Molecular Microbiology 36(5), 1006-1014.

Kovats, R.S., Edwards, S.J., Hajat, S., Armstrong, B.G., Ebi, K.L., Menne, B. 2004. The effect of temperature on food poisoning: a time-series analysis of salmonellosis in ten European countries. Epidemiol. Infect. 132, 443–453.

Kuchenmüller, T., Hird, S., Stein, C., Kramarz, P., Nanda, A., Havelaar, A.H. 2009. Estimating the global burden of foodborne diseases – a collaborative effort. Eurosurveillance 14(18), 1-4.

Lake, R.J, Cressey, P.J., Campbell, D.M., Oakley, E. 2010. Risk Ranking for Foodborne Microbial Hazards in New Zealand: Burden of Disease Estimates. Risk Analysis 30(5), 743-752.

Lawson, A.B. 2009. Bayesian disease mapping: hierarchical modelling in spatial epidemiology. New York:Chapman and Hall/CRC Interdisciplinary Statistics. 368p. Ill.

Majowicz, S.E., Musto, J., Scallan, E., Angulo, F.J., Kirk, M., 4,5,Sarah J. O'Brien, S.J., Jones, T.F., Fazil, A., Hoekstra, R.M. 2010. The Global Burden of Nontyphoidal *Salmonella* Gastroenteritis. Clin Infect Dis. 50(6), 882-889.

Mangen, M.J., Batz, M.B., Käsbohrer, A.K., Hald, T., Morris Jr., J.G., Taylor, M, Havelaar, A.H. 2010. Integrated Approaches for the Public Health Prioritization of Foodborne and Zoonotic Pathogens. Risk Analysis 30(5), 782-797.

McBride, G., Meleason, M., Skelly, C., Lake, R., VanDerLodgt P., Collins, R. 2005. Preliminary Relative Risk Assessment for Campylobacter Exposure in New Zealand. NIWA Client Report HAM2005-094.

Mead, P. S., Slutsker, L., Dietz, V., & McCaig, L. F. 2000. Food-related illness and death in the United States. Journal of Environmental Health, 62(7), 9-18.

Miller, S.I., Hohmann, E.L., Pegues, D.A., 1995. *Salmonella* (including *Salmonella typhi*). In: Mandell, G.L., Bennett, J.E., Dolin, R. (Eds.). Principles and practice of infectious diseases. Churchill Livingstone, New York, pp. 2013-2033.

Mølbak, K., Neimann, J. 2002. Risk Factors for Sporadic Infection with *Salmonella* Enteritidis, Denmark, 1997–1999. Am. J. Epidemiol 156(7), 654-661.

Mølbak, K., Olsen, J.E. and Wegener. H.C. (2006). Salmonella infections. In: Foodborne infections and intoxications. Third edition. Eds.Rieman, H.P and D.O. Cliver. School of Veterinary Medicine, University of California, Davis. Academic press, Elsevier.

Müllner, P., Jones, G., Noble, A., Spencer, S.E. F., Hathaway, S., French, N.P. 2009. Source Attribution of Food-Borne Zoonoses in New Zealand: A Modified Hald Model. Risk Analysis 29(7), 970-984.

Müllner, P., S. E. F. Spencer, D. Wilson, G. Jones, A. D. Noble, A. C. Midwinter, J. Collins-Emerson, P. Carter, S. Hathaway, and N. P. French. 2009. Assigning the source of human campylobacteriosis in New Zealand: a comparative genetic and epidemiological approach. Infect. Genet. Evol. 9:1311–1319.

Müllner, P., Collins-Emerson, J.M., Midwinter, A.C, Carter,P, Spencer,S.E.F., van der Logt, P., Hathaway, S., French, N.P. 2010. Molecular Epidemiology of Campylobacter jejuni in a Geographically Isolated Country with a Uniquely Structured Poultry Industry. Appl. Environ. Microbiol. 76(7):2145-2154.

Newell, D.G., Koopmans, M., Verhoef, L., Duizer, E., Aidara-Kane, A. Sprong, H., Opsteegh, M., Langelaar, M., Threfall, J., Scheutz, F., van der Giessen, J., Kruse, H. 2010. Food-borne diseases — The challenges of 20 years ago still persist while new ones continue to emerge. International Journal of Food Microbiology 139, S3–S15.

Nygaard, K., de Jong, B., Guerin, P.J., Andersson, Y., Olsson, A., Giesecke, J. 2004. Emergence of new *Salmonella* Enteritidis phage types in Europe? Surveillance of infections in returning travelers. BMC Medicine, 2:32.

O'Reilly, C.E., Bowen, A.B., Perez, N.E., Sarisky, J.P., Shepherd, C.A., Miller, M.D., Hubbard, B.C., Herring, M., Buchanan, S.D., Fitzgerald, C.C., Hill, V., Arrowood, M.J., Xiao, L.X., Hoekstra, R.M., Mintz, E.D., Lynch, M.F., 2007. A waterborne outbreak of gastroenteritis with multiple etiologies among resort island visitors and residents: Ohio, 2004. Clin. Infect. Dis. 44, 506-512.

Pires, S.M. 2009. Attributing human salmonellosis and campylobacteriosis to food, animal and environmental sources. PhD Thesis. University of Copenhagen. ISBN 978-87-7611-311-7.

Pires, S.M., Evers, E.G., van Pelt, W., Ayers, T., Scallan, E., Angulo, F. J, Havelaar, A., Hald, T. 2009. Attributing the Human Disease Burden of Foodborne Infections to Specific Sources. Foodborne Pathogens and Disease 6(4), 417-424.

Pires, S.M., Hald, T. 2010. Assessing the differences in public-health impact of *Salmonella* subtypes using a Bayesian microbial subtyping approach for source attribution. Foodborne Pathogens and Diseases, 7(2): 143-151.

Pires, S.M., Vigre, H., Makela, P., Hald, T. 2010. Using outbreak data for source attribution of human salmonellosis and campylobacteriosis in Europe. Foodborne Pathogens and Disease, 7(11): 1351-1361.

Pires, S.M., De Knegt, L.V., Hald, T. 2011a. Estimation of the relative contribution of different food and animal sources to human *Salmonella* infections in the European Union. Scientific Technical Report submitted to EFSA. EFSA-Q-2010-00685. Available at http://www.efsa.europa.eu/en/supporting/doc/184e.pdf.

Pires, S.M., Toyofuku, H. Kasuga, F., Hald, T., 2011b. Attributing foodborne disease in Japan using outbreak data. International Association for Food protection Annual Meeting, Milwaukee, Winsconsin, August July 31st- August 3rd, 2011.

Pires, S.M., Vieira, A.R., Perez, E., Lo Fo Wong, D.M.A., Hald, T. 2012. Attributing human foodborne illness to food sources and water in Latin America and the Caribbean using data from outbreak investigations. International Journal of Food Microbiology 152(2), 129-138.

Ravel, A., Greig, J., Tinga, C., Todd, E., Campbell, G., Cassidy, M., Marshall B., Pollari, F. 2009. Exploring historical Canadian foodborne outbreak data sets for human illness attribution. Journal of Food Protection 72(9):1963-76.

Rosef, O., Kapperud, G., Lauwers, S., Gondrosen, B. 1985. Serotyping of Campylobacter jejuni, Campylobacter coli and Campylobacter laridis from domestic and wild animals. Appl. Environ. Microbiol. 49:1507–1510.

Scallan, E., Hoekstra, R.M, Angulo, F.J., Tauxe, R.V., Widdowson, A., Roy, S.L., Jones, J.L., Griffin, P.M. 2011a. Foodborne Illness Acquired in the United States—Major Pathogens. Emerging Infectious Diseases 17(1), 7-15.

Scallan, E., Griffin, P.M., Angulo, F.J., Tauxe, R.V., Hoekstra, R.M., 2011b. Foodborne Illness Acquired in the United States—Unspecificed agents. Emerging Infectious Diseases 17(1), 16-22.

Semenza, J.C., Herbst, S., Rechenburg, A., Suk, J.E., Hoeser, C., Schreiber, C., Kistemann, T. 2012. Climate Change Impact Assessment of Food- and Waterborne Diseases. Critical Reviews in Environmental Science and Technology

Senior, K. 2009. Estimating the global burden of foodborne diseases. Lancet Infect. Dis. 9, 80-81.

Sharma S. 1996. Cluster analysis. In: Applied multivariate techniques. John Wiley&Sons, New York:USA. 491pp. il.y 42(8), 857-890.

Smith, A.F.M., Gelfand, A.E. 1992. Bayesian statistics without tears: a sampling-resampling perspective. The American Statistician, Vol. 46, No. 2. (May, 1992), pp. 84-88.

Tauxe, R.V., Doyle, M.P., Kuchenmüller, T., Schlundt, J., Stein, C. 2010. Evolving public health approaches to the global challenge of foodborne infections. International Journal of Food Microbiology 139, S16–S28.

TESSy. The European Surveillance System. Acessed on June 15, 2011. Available online at http://www.ecdc.europa.eu/en/activities/surveillance/TESSy/Pages/TESSy.aspx).

Tindall, B. J., Grimont, P.A.D., Garrity, G.M., Euzéby, J.P. 2005. Nomenclature and taxonomy of the genus *Salmonella*. International Journal of Systematic and Evolutionary Microbiology 55, 521–524.

Torpdahl, M., G. Sørensen, B.-A. Lindstedt, E. M. Nielsen, 2007. Tandem repeat analysis for surveillance of human *Salmonella* Typhimurium infections. Emerg. Infect. Dis. 13: 388-395.

Toyofuku, H., Pires, S., Hald, T. 2011. *Salmonella* Source Attribution in Japan by a Microbiological Subtyping Approach. Ecohealth 7 Issue Suppl. 1, S22-S23.

Uzzau, S., Brown, D.J., Wallis, T., Rubino, S., Leori, G., Bernard, S., Casadesu, J., Platt, D. J., Olsen, J. E. 2000. Host adapted serotypes of *Salmonella* enterica. Epidemiol. Infect. 125, 229-255.

Valkenburgh, S., Oosterom,R, Van Stenvers, O., Aalten, M., Braks, M., Schimmer, B., Giessen, A., Van Pelt, W., and Langelaar, M. 2007. Zoonoses and Zoonotic Agents in Humans, Food, Animals and Feed in The Netherlands 2003–2006. RIVM rapportnummer: 330152001. ISBN-13, 978-90-6960-184-7.

van Duijkeren, E., Wannet, W.J.B., Houwers, D.J., van Pelt, W. Serotype and Phage Type Distribution of *Salmonella* Strains Isolated from Humans, Cattle, Pigs, and Chickens in The Netherlands from 1984 to 2001. J. Clin. Microbiol. 40(11), 3980-3985.

van Pelt W, Van De Giessen AW, Van Leeuwen WJ, et al. Oorsprong, omvang en kosten van humane salmonellose. Deel 1. Oorsprong van human salmonellose met betrekking tot varken, rund, kip, ei en overige bronnen. Infectiezikten Bulletin 1999;10:240–243.

van Valkenhoef, G., Lu, G., de Brock, B., Hillege, H., Ades, A.E., Welton, N.J. 2012. Research Synthesis Methods 3(4):285–299.

Voetsch, A.C., Gilder, T.J.Van, Angulo, F. J., Farley, M.M., Shallow, S., Marcus, R., Cieslak, P.R., Deneen, V.C., Tauxe, R.V. 2004. FoodNet Estimate of the Burden of Illness Caused by Nontyphoidal *Salmonella* Infections in the United States. Clin Infect Dis. 38(Supplement 3), S127-S134.

Vose, D., Koupeev, T. and Mintiens, K. 2011. Scientific/Technical report submitted to EFSA on a quantitative microbial risk assessment of Salmonella spp. in broiler (Gallus gallus) meat production. Report to contract CT/EFSA/BIOHAZ/2010/02. Vose Consulting (US) LLC, 81 pp.

Wegener, H.C., Hald, T., Lo Fo Wong, D.M.A., Madsen, M. Korsgaard, H., Bager, F., Gerner-Smidt, P., Mølbæk, K. 2003. *Salmonella* control programs in Denmark. Emerging Infectious Diseases 9(7), 774-780.

Wegener, H. 2010. Danish initiatives to improve the safety of meat products. Meat Science 84:276-283.

Weill, FX., et al. 2006. Ciprofloxacin-resistant *Salmonella* Kentucky in Travelers. Emerging Infectious Diseases.12(10): 1611–1612.

Whalström, H, Andersson, Y, Plym-Forshell, L, Pires, SM. 2011. Source attribution of human *Salmonella* cases in Sweden. Epidemiology and Infection 139:1246-1253.

Wheeler, J.G., Sethi, D., Cowden, J.M., Wall, P.G., Rodrigues, L.C., Tompkins, D.S., Hudson, M.J., Roderick, P.J. 1999. Study of infectious intestinal disease in England: rates in the community, presenting to general practice, and reported to national surveillance. The Infectious Intestinal Disease Study Executive. BMJ 318(7190), 1046-1050.

Wilson, D.J., Gabriel, E., Leatherbarrow, A.J.H., Cheesbrough, J., Gee, S., Bolton, E., Fox, A., Fearnhead, P., Hart, C.A., Diggle, P.J. 2008. Tracing the source of campylobacteriosis. PLOS Genetics 4(9), 1-9.

WHOCC-Salm, 2007. Antigenic formulae of the *Salmonella* serovars. 9th Edition. WHO Collaborating Centre for Reference and Research on *Salmonella*. Institut Pasteur:Paris. 166p.

World Health Organization. 2002. Methods for Foodborne Disease Surveillance in Selected Sites. Report of a WHO consultations. WHO/CDS/CSR/EPH/2002.22.

World Health Organization 2007. The world health report 2007. Geneva, Switzerland: World Health Organization.

World Health Organization. 2009. Second formal meeting of the Foodborne Disease Burden Epidemiology Reference Group (FERG): appraising the Evidence and Reviewing Initial Results. ISBN 978.92.4.159849.1

MANUSCRIPTS

Manuscript I: Using surveillance and monitoring data of different origin in a *Salmonella* source attribution model: a European Union example with common challenges and proposed solutions

Manuscript II: Attributing foodborne salmonellosis in humans to animal reservoirs in the European Union using a multi-country stochastic model

Manuscript III: Sources of human salmonellosis in Denmark: comparing the results of the Danish *Salmonella* source account model with a source attribution model developed at EU level

Manuscript I

Using surveillance and monitoring data of different origins in a *Salmonella* source attribution model: a European Union example with challenges and proposed solutions

L.V. DE KNEGT[(1)], S.M. PIRES[(1)], T. HALD[(1)]

[(1)],National Food Institute, Division of Epidemiology and Microbial Genomics

Technical University of Denmark

Author responsible for manuscript and reprint correspondence: L.V. DE KNEGT

Mailing address:

Technical University of Denmark, National Food Institute

Division of Epidemiology and Microbial Genomics

Mørkhøj Bygade 19, Building H, 2860 Søborg, Denmark

Email: ledkn@food.dtu.dk

Running head: European data for source attribution

Submitted to Epidemiology and Infection

Summary

Microbial subtyping approaches are commonly used for source attribution of human salmonellosis. Such methods require data on *Salmonella* in animals and humans, outbreaks, infection abroad and amounts of food available for consumption. A source attribution model was applied to 24 European countries, requiring special data management to produce a standardized dataset. Salmonellosis data in animals and humans were obtained from datasets provided by European Food Safety Authority. The amount of food available for consumption was calculated based on production and trade data. Limitations included different types of underreporting, non-participation in prevalence studies, and non-availability of trade data. Cases without travel information were assumed to be domestic; non-subtyped human or animal records were re-identified according to proportions observed in reference sources; missing trade information was estimated based on previous years. The resulting dataset included data on 24 serovars in humans, broilers, laying hens, pigs and turkeys in 24 countries.

INTRODUCTION

Unsafe food is related to several kinds of diseases, ranging from diarrhoeal syndromes to various forms of cancer [1,2]. Although the true burden of foodborne diseases is currently unknown, it is considered that is has increased in the last decades, as the growth of global population was accompanied by an increase of consumption of animal products, and the more intensive production methods required to supply it [1,3]. In 2005, it was estimated that food- or waterborne diarrhoeal diseases were responsible for 2.2 million deaths per year worldwide, 1.9 million of which were children [2].

Salmonella spp. is one of the most common and widely distributed foodborne pathogens in the European Union (EU), with a total of 108.614 laboratory-confirmed cases reported by 27 Member States (MS) in 2009. Athough its relative importance has been decreasing since 2006, *S*.Enteritidis is still the main reported serovar (52.3% of cases), followed by *S*.Typhimurium (23.3%). However, a wide range of others frequently cause disease in humans and thus are of public health significance [4].

Since 2003, efforts have been made in the EU to standardize the reporting of pathogens and diseases in humans and animals. These included the conduction of studies to estimate the MS-level baseline prevalence of *Salmonella* in animals of the food chain [5,6,7,8], and setting targets to reduce it. Other efforts were the harmonization of the monitoring of *Salmonella* in laying hens [9], broilers [10] and turkeys [11], the last two implemented after the activities described in this manuscript. Those actions are expected to have an impact on the contribution of different food-animals to human salmonellosis in all individual MSs, but until 2010, this information had not been assessed.

Identifying which foods are more frequently implicated in the transmission of an illness is a crucial step on the prioritization of control measures [12], and a variety of methods to attribute foodborne pathogens to specific sources are available, including approaches based on analysis of data from microbiological and epidemiological studies, intervention studies, and expert elicitations [13]. Source attribution (SA) methods present different advantages and limitations, and their applicability depends on the pathogen in question and on the data available to address a specific public health question [13,14].

Several *Salmonella* SA studies based on microbial subtyping have been conducted in EU countries with well-established public health and animal surveillance systems [15,16,17,18]. At the EU-level, an analysis of outbreak data for SA of salmonellosis was conducted [19], and its results suggested regional differences in the relative importance of food sources for disease, but also reflected the variability in the effectiveness of implemented surveillance systems and quality of data in different countries. For that reason, no direct comparison of the public health impact of food sources between EU countries or regions was possible.

The principle of source attribution by microbial subtyping is to compare the occurrence of subtypes in animals or food sources with the same subtypes in humans, provided that subtypes are heterogeneously distributed among the sources. Human infections caused by source-specific subtypes are attributed to the corresponding sources. Infections caused by subtypes found in several reservoirs are distributed relatively to the prevalence of the specific types. This approach requires an integrated foodborne disease surveillance programme that collects isolates from the major food-animal reservoirs of foodborne diseases, as well as information on sporadic human cases, outbreaks and travel-related cases [15].

Based on the SA studies reviewed [15,16,17,18], the "perfect" dataset would include 1) the number of reported salmonellosis cases in humans, originating from a nationally representative surveillance system in which cases are all confirmed by laboratory and subtyped to an appropriate discriminatory level; 2) information on whether the person reported had been travelling abroad up to seven days prior to symptoms onset; 3) number of outbreak cases and identified outbreak sources; 4) prevalence of *Salmonella* subtypes characterized by the same subtyping methods as applied to human isolates and representing all major sources of human salmonellosis in Europe and 5) the amount of an animal product originating from a country which is ingested by consumers in another country. Phage type data with further differentiation based on antimicrobial resistance profiling is currently considered the ideal level of subtyping for those models, as it allows better differentiation of common subtypes (e.g. *S.* Enteritidis and *S.* Typhimurium) among similar sources, when compared to using serovars [20].

The present paper describes the data obtained from different sources available in 2010 to be used in an EU-wide SA model based on microbial subtyping, as well as the data management steps taken to produce a

homogenous dataset containing *Salmonella* serovar information from humans and animal reservoirs used for food production. Limitations to the data available are presented, along with the solutions applied to solve them.

METHODS

Data sources

The European Surveillance System (TESSy): TESSy is a system for collection, validation, analysis and dissemination of data from 27 EU MSs and three European Economic Area (EEA) countries, functioning since 2008 [21]. Countries report their data on communicable diseases to the system, which is administered by the European Centre for Disease Prevention and Control (ECDC). The system also records information on outbreaks and possibility of infection during international travel. The specific reporting of *Salmonella* serovars is only mandatory for *S*.Enteritidis and *S*.Typhimurium [21], meaning that the reporting of other serovars and further subtyping levels is only done on a voluntary basis. Data reported prior to 2008 are also available, since TESSy replaced the data collection systems for the Data Surveillance Network, which collected national data individually [21].

The European Union Summary Report on Trends and Sources of Zoonoses, Zoonotic Agents and Foodborne Outbreaks (EUSR): The report has been prepared by the European Food Safety Authority (EFSA) since 2004 in collaboration with ECDC. Data on the occurrence of zoonoses and zoonotic agents in animals, foodstuffs and animal feed is reported annually by MSs to EFSA and summarized in the EUSR. Serovar reporting in the animal data reported to the EUSR follow the same requirements as described for humans.

Baseline studies on the prevalence of Salmonella in animal populations in the European Union (BS): In order to provide the scientific basis for setting prevalence targets for reduction of *Salmonella* in commercial and breeding farms, EU-wide studies on the baseline prevalence of *Salmonella* were conducted focusing on laying hens (2004-2005) [5], broiler flocks (2005-2006) [8], slaughter pigs (2006-2007) [6], fattening and breeding turkeys (2006-2007) [7], broiler carcasses (2008) [22] and breeder pigs (2008) [23]. The studies took place during a four-year period, and varied in MS participation due to the addition of new members to the EU in

2004 and 2007, and to the occasional participation of EEA countries. However, they still constitute the most uniformly collected and analyzed data on *Salmonella* at EU level, allowing valid comparisons among MSs.

The statistical office of the European Union (EUROSTAT) [24]: EUROSTAT was established in 1953 to provide the European Union with statistics at European level that enable comparisons between countries and regions. It collects data on the value and quantity of food and slaughter animals traded between EU MSs and by EU MSs with third countries. European Community legislation ensures that the statistics provided to EUROSTAT by the MSs are based on legal texts and on harmonized definitions and procedures [24]. However, data availability varies depending on country and products selected, since the information is provided directly by MS, being subject to variations in national focus and cultural differences.

Data were stored and analyzed in SAS Enterprise Guide, SAS Institute, SAS/STAT® User's Guide, Version 8, Cary, NC: SAS Institute Inc., 1999.

Reported cases of human salmonellosis

Data on the number and serovar distribution of human cases reported to TESSy from 2007 to 2009 were extracted on 6th of July 2010 and provided by ECDC through EFSA. The total number of reported cases includes sporadic, travel and outbreak-related infections. MSs for which the level of serovar detailing was insufficient for source attribution were requested to provide additional data if available. Such national datasets were provided by Poland and Portugal. The MSs providing data on sporadic cases and outbreaks are summarized in Table 1.

Challenge 1: Underreporting

One issue arising from the use of surveillance data is the underreporting of cases, which can happen in all steps of the surveillance process [25]. It is generally understood that the real (and generally unknown) number of illnesses occurring in the population is larger than the number of cases that actually get reported in the surveillance system, which is explained by the percentage of: 1) cases who seek medical care; 2) cases which are asked for clinical specimens and actually provide them; 3) specimens which are tested; 4) sensitivity of the laboratory tests used and 5) positive results which are reported [25]. Therefore, it is accepted

that the true burden of human salmonellosis may be considerably larger than the officially reported incidence. The level of underreporting is expected to vary between countries, depending on differences in organization and effectiveness of local surveillance systems [21,26].

Proposed solution: In 2012, Havelaar et al.[21,26] used data from a Swedish travel database and the *Salmonella* incidence from a Dutch population-based study to estimate a set of multipliers for correction of underreporting in 31 European countries. The multipliers were estimated based on the proportion of cases of salmonellosis which were reported in Sweden upon returning from the Netherlands, and represent an estimation of the number of cases that should have been reported for each case that entered the system. It is expected that the use of these multipliers have an impact on the most important sources estimated at EU-level. As the adjustment for underreporting is only done after the attribution process, the corrected numbers are not shown here, but can be found in de Knegt [20].

Challenge 2: Incomplete travel-related information

Travel information, derived from data reported as "probable country of infection" was reported as "imported", "not-imported" or "unknown location of origin". The proportion of travelers and the amount of information provided varied among MSs; in Sweden and Finland, travel-related cases corresponded to 77% and 82% of the total, while other countries (nine in 2009) reported 100% of cases as "unknown travel history".

Proposed solution: The Hald model and its adaptations [15,16] use the observed proportion of travel cases that were properly discriminated to redistribute cases with no information to the "travel-related" and "domestic" categories; the same approach could potentially be used to estimate extra travelers in the EU model. In case there is not enough information available for the redistribution, cases which did not specifically report a travel history should be considered as domestic,

Challenge 3: incomplete or missing serovar identification

Expected situations in which serovar identification is missing or incomplete can be summarized as: a) classification only up to genus or species level, such as *Salmonella* spp, or *Salmonella enterica*; b)

classification up to subspecies level, such as *Salmonella enterica enterica* or *Salmonella ente*rica Subspecies I; c) classification using groups based on the O-antigen both by the old nomenclature, like groups B, C1-C2 or E4, or the new one, such as serogroups O:4, O:7 or O:33; d) aggregated data, where the main serovars were specified, and the remaining were grouped as "Others"; e) cases where the serovar field is simply blank or filled with "unknown".

Proposed solutions: isolates not classified up to serovar level or data reported in aggregated form should be reassigned to specific serovars according to proportions observed in previous studies, in the same dataset or in other references, depending on the availability of data in each case.

Isolates identified up to genus or species level, blank or filled with "unknown" should be reassigned to all serovars observed in the country. (e.g.: if *S.* Enteritidis accounts for 60% of all serotyped isolates from human cases in a country, and 10 isolates in the same country receive one of the denominations mentioned, six of them must be reassigned to *S.* Enteritidis). Isolates identified up to subspecies level should likewise be reassigned to all serovars in the country, but with proportions calculated using only isolates of *S. enterica enterica* as total.

Isolates classified as serogroups should be distributed among serovars pertaining to those groups, in accordance with the Kauffman-White-Le Minor Scheme 9th edition [27] (e.g., if *S.*Typhimurium accounts for 40% of all isolates in the country, but for 80% of units from serovars belonging to Group B, and 10 isolates are only identified as "*Salmonella* Group B", eight of those must be reassigned to *S.*Typhimurium).

Isolates classified as "Others" are assumed to belong to serovars not described in the current dataset, but nonetheless present in the country. In this case, the reference used for reassignment proportions is the World Health Organization Global Foodborne Infections Network (GFN) Country Databank (CDB) [28], which contains the 15 most commonly identified *Salmonella* serovars among human and non-human sources in 84 countries (e.g.: in the original TESSy data, a country reports 30 isolates: 10 *S.*Enteritidis, 10 *S.*Typhimurium and 10 "Others". The GFN CDB shows 80% *S.*Enteritids, 10% *S.*Typhimurium, 7% *S.* Infantis and 3% *S.*Hadar for this country, so, according to this reference, *S.*Infantis and *S.*Hadar correspond to 70% and 30% of the non-described serovars. The 10 isolates should then be redistributed as seven *S.*Infantis and three

S.Hadar, as it is assumed that no *S*.Typhimurium or *S*.Enteritidis isolates are included in the group of "others").

Challenge 4: underreporting and incomplete identification of serovars in outbreak data

For outbreaks of foodborne salmonellosis, the same datasets used for the EUSRs 2007-2009 [4,29,30] were provided by EFSA. Not all countries report outbreak cases, and not all reported cases are reported with complete serovar information.

Proposed solutions: the same underreporting multipliers used for sporadic cases cannot be applied to outbreaks, as it is assumed that outbreaks have a higher probability of being reported. Based on that, countries which report sporadic cases but no outbreak cases are assumed as not having any foodborne *Salmonella* outbreaks in the period. Outbreak-related cases for which a serovar is not fully identified should be reassigned using the proportions observed in the same outbreak dataset, as some serovars may be more prone to generate outbreaks than others [16], and thus the proportions observed in reported sporadic cases may not apply.

<u>*Salmonella* in livestock and food</u>

Challenge 5: Data available may not be representative of all MSs and animal sources

Data from the EU BS on the prevalence of *Salmonella* the sources were the preferred data source. Due to admission of new MSs to the EU and to the fact that participation in the BSs is voluntary, it is expected that BS data is not available for all sources in all MSs. However, these datasets were still considered the most representative of the given reservoirs, since no harmonized EU monitoring in pigs and turkeys was currently in place, and the broiler carcass study was considered to provide more recent data than BS on broiler flocks, with a better detailing of the serovar distribution when compared to the existing EU monitoring data. The laying hens BS was conducted between 2004 and 2005 [5], and it is expected that the implementation of the harmonized monitoring [9] has resulted in significant changes in the *Salmonella* serovar prevalences in this reservoir in many MSs. No data from BS or EU-harmonized monitoring exist for cattle.

Proposed solutions: in order to use the most recent data possible, data which is missing from BSs should be supplied with surveillance and monitoring data found in the EUSR. When not enough surveillance or

monitoring data at herd/flock level are available for a source or MS, slaughter samples should be surveyed and their quality as substitutes assessed.

Challenge 6: incomplete or missing serovar identification

In the laying hens data, in addition to isolates with non-identified or partially identified serovars, many countries only report a reduced list of serovars and a group of "Others", as *S.*Enteritidis and *S.*Typhimurium are the only two serovars for which specific reporting is mandatory [21]. For BS data, no reference for reassigning of serogroups or incomplete serovar identification is available.

Proposed solutions: proportions found in the laying hens BS [5] should be used for re-allocation of laying hen units. In datasets where there are no records identified as "Others", units should be redistributed according to the proportions found among properly identified serovars in the same dataset. The criteria for reassigning non-identified or partially identified serovars should be the same as for the human data.

Food production and trade data

Food production data were derived by EFSA from the EUROSTAT databases on production and on slaughtered animals for food consumption [24]. Consumption calculations were based on trade data. This was done so the attribution model can account for the amount of food present in a given country which originated from other countries and use the country- and food-specific serovar prevalences for the attribution [20]. The domestically produced amount available for consumption of a food source in a MS was estimated as Domestic Production minus Export, whereas the amount of imported food available for consumption in MS A originating from MS B was estimated as Import minus Re-export (when re-export was relevant), thus making it necessary the use of production data, as well as country-to-country imports and exports. For this study, extra-EU food trade was not considered [20].

Challenge 7: Missing data

Information on poultry for meat production was not available for Belgium in 2007 and 2008. Egg production data were lacking for several countries, and data for most food sources and most years were missing in

Cyprus. Data on the export of the food sources to other MSs included in this study were available for all considered countries, with the exception of the amount of eggs exported from Cyprus.

Proposed solutions: Missing data on annual quantities of poultry meat products sold per MS, with differentiation between boilers, turkeys and other poultry species should be obtained from the 2009 Annual Report of the Association of Poultry Processors and Poultry Trade in the EU Countries [31]. For all sources, countries with information missing for a year should have the missing value estimated based on the percentage of increase or decrease between available years; when data from only one year is available, that value will be used as surrogate for the missing years.

Challenge 8: Negative estimated amounts available for consumption

Due to differences among numbers reported in the production, imports and exports datasets, the operation to calculate the amount of a food source available for consumption in a country in some cases results in negative numbers, meaning the volume exported is larger than the domestic production.

Proposed solution: In order to ensure that MSs will still have nationally produced food available in their own country, it re-exporting of imported products should be considered possible.

Challenge 9: Validation of the estimation of consumption data based on trade data.

The underlying assumptions for this estimation were that EUROSTAT data were complete and consistent and that all the food available for consumption is actually consumed, in a way that these data reflected the real flow of foodstuffs and consequent exposure in the countries. According to a quality assessment performed by EFSA [32], the information recorded in those datasets does not fully support those assumptions. This assessment showed the existence and non-reporting of triangular trade, mis-classification of food products and problems in the conversion of currency/weight units. Also, we expect that in several situations, data for missing years needs to be estimated or supplied with surrogate data (e.g. AVEC data), resulting in a highly manipulated dataset that may not represent reality.

Proposed solution: the data management can be validated by comparing the resulting consumption dataset with consumption data available from the WHO Global Environment Monitoring System Food Consumption

Cluster Diets [33]. The WHO data is available in grams/person/day, so the estimated data should be converted to the same unit. As the WHO data only offers the broad category "poultry", broilers and turkeys should be summed. Relative proportions of consumption of poultry, pork and eggs must be calculated, so a Proportional Similarity Index (PSI /Czekanowsky index) can be used to compare those proportions between the two groups in each country. The PSI is an estimate of the area of intersection between two frequency distributions [34], and is calculated as

$$PSI= 1-0.5*\sum|p1-q1| = \sum \min(p_1,q_1)$$

It is traditionally used for calculating niche overlap and resource availability in population ecology [35] or proportions of identified bacterial strains in epidemiology [36,37], but here it was considered that each of the relative proportions among the three sources corresponds to the area under a probability curve, and so the same measure can be applied. A PSI of 1 means a complete overlap, or 100% similarity. An "overall PSI" for the whole dataset was calculated by averaging the country PSI values.

RESULTS

The availability of data in all surveyed countries is shown in Table 1. A list of 25 serovars was selected to be addressed further, based on their occurrence in humans and animals (Table 2).

Human data

The percentage of records that had incomplete identification and had to be reassigned varied from zero in Portugal to 84% in Romania (Table 3). The most common reason for reassignment were records reported in aggregated form, i.e., with several serovars under a group named "Others", and the second were isolates reported as "Unknown", followed by isolates only classified as serogroups (Table 3). Besides the predicted identification problems, a specific issue regarding *S*. Typhimurium was found: one of the defining characteristics of *S*.Typhimurium is presenting two phases of H-antigens: "i" and "1,2", which is why the antigenic formula for this serovar is written as "1,4,[5],12:i:1,2", with the two mentioned phases seen at the end [27]. However, variants that lack either the first- or the second-phase H antigen have been described, and reported by some countries as "1,4,[5],12:i:-", "1,4,[5],12:-:1,2" or "1,4,[5],12:-:-". *S*.Typhimurium-like

variants with only one phase of the H or I antigens are referred to as "*S*.Typhimurium-like strains" or "Monophasic *S*.Typhimurium" [38]. For our purposes, isolates identified by those formulas in the datasets were reassigned to *S*.Typhimurium, which is supported by an EFSA Biohazard Panel assessment [38]. The aphasic antigenic formula "1,4,[5],12:-:-" was not reassigned, as it could belong to several serovars in group O:4.

Regarding outbreak data, Bulgaria, Cyprus, Greece, Italy, Luxembourg, Malta and the United Kingdom did not report any cases. Nearly 47% of outbreak cases reported by France had to be reassigned, as the isolates were reported as "*Salmonella* spp". For Latvia, the proportion was 39% (Table 4). Switzerland reported outbreaks, but no sporadic cases (Table 1).

Travel information (Table 5) was reported as "Unknown" for 100% of isolates in France, Romania and Slovenia. Full travel information was provided by Austria, Belgium, the Czech Republic, Estonia, Spain, Hungary, the Netherlands and Slovakia. The remaining MSs had variable proportions of cases reported as "Travel-related", "Domestic" or "Unknown". As a result, the proposed "informed redistribution" was not possible, as a large number of countries did not report any travel cases. As a consequence, all records with missing or unknown travel information from countries with serovar detailing of sporadic cases were considered domestically acquired in the reporting country.

Figure 1 shows, , the relative occurrence of the 11 serovars most frequently found simultaneously in humans and animals in the last five years in sporadic and outbreak cases[4,30]. *S*.Enteritidis and *S*.Typhimurium were the serovars most frequently observed in sporadic human cases, together with *S*.Infantis, *S*.Newport, *S*.Kentucky, *S*.Virchow, *S*.Derby and *S*.Agona. The most common serovars observed in outbreaks were *S*.Enteritidis and *S*.Typhimurium. As expected, outbreaks may be associated with serovars not normally found in the country. That is particularly true in countries with a small number of sporadic cases and a good level of control of *Salmonella* in domestic products, like Finland or Sweden.

Animal-food data

Data was available from 28 countries (see Table 1 for data origin in each reservoir). Data for laying hens were obtained from the EUSR 2008 [30], which was the first year of EU-harmonized reporting for this

reservoir, being preferred over the BS data. Greece did not participate in the broiler carcasses study [22], being supplied with data from the broiler flocks BS [8]. Malta and Romania did not participate in the study on slaughter pigs [6], and no surrogate data was available for those countries. For turkeys, BS data from fattening flocks were chosen over breeding flocks [7], with the exception of Estonia, Latvia, Luxembourg and Romania, which were not part of the study. Data for those countries were provided by EUSR data from 2006 and 2008 [30,39], but no surrogate data was available for Romania. Non-harmonized surveillance data on cattle, including carcass samples at slaughter, was retrieved from the EUSR 2007, 2008 and 2009 [4,29,30] with 2009 data being preferred to the other years. Cattle data for France was retrieved form a PhD thesis [40]. For this reservoir, no data from Cyprus or Malta were identified, and for some countries only a single year of data was available. In the resulting datasets, Belgium and the United Kingdom only reported positive samples for cattle, resulting in 100% positivity in those countries. Small samples were also observed for broilers in Luxembourg, laying hens in Lithuania and Luxembourg, and turkeys in Estonia, Luxembourg and Latvia. Total samples submitted and total positive per country are summarized in Table 6. The amount and percentage of reassigned records among the total positives are shown in Table 7.

Serovar predominance varied between countries in all animal sources. Overall, considering the relative occurrence of serovars and number of countries in which they predominated, *S*. Infantis and *S*. Enteritids were the main serovars observed in broilers, *S*. Typhimurium and *S*. Derby in pigs, *S*. Typimhurium, *S*. Bredeney and *S*. Hadar in turkeys, *S*. Enteritidis and *S*. Infantis in layers. S. Dublin and S. Tpyhimurium were the main serovars in cattle, but the data was considered too heterogeneous and frail to be considered representative. The top-ten serovars for broilers, pigs, turkeys and layers and relative proportion graphs for the selected serovars can be found in de Knegt [20].

Trade and consumption data

Availability of data on the annual quantities of poultry, pork and bovine meat and eggs produced varied per year and per MS. All MSs reported imports from other MSs for all food products in the study period. The resulting surrogate consumption dataset was considered valid, as shown by the results of the data validation by comparison with GEMS data (Table 8). The individual PSI values were higher than 0.8 in all countries

except for one, indicating more than 80% similarity between the estimated data and the reference values. The one exception was Cyprus, with only 42% similarity, which is expected to have an impact on the attribution estimates for this country. Still, the overall PSI was 0.91, indicating that the dataset as a whole can be used without considerable bias.

Final dataset for the source attribution model

Based on data availability and quality, 24 countries were included in model: Austria, Belgium, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden, The Netherlands and the United Kingdom. Countries initially analyzed and excluded from the final dataset were Bulgaria, which provided 100% of human cases without serovar detailing; Romania, which only participated in one BS, did not have enough surrogate data to be retrieved from the EUSR, and reported 84% of cases without serovar information; Norway and Switzerland, which do not report to EUROSTAT, the latter also not reporting to TESSy.

Based on the availability of EU-wide homogeneous data or with at least good-quality surrogates, food-animal sources included were broilers, pigs, turkeys and laying hens (as the animal reservoirs for chicken meat, pork, turkey meat and eggs). Due to better completeness and availability, the resulting trade data from 2009 was used as consumption data for those sources. Data from the cattle reservoir were in general poor and for some MSs consisting of clinical isolates only. Efforts to improve the dataset by using herd information from 2007-2008 or slaughterhouse carcass samples did not prove sufficient to obtain a representative dataset for this source.

Twenty-two serovars were selected to be specifically addressed, based on their presence and importance in humans and in the main animal reservoirs: *S.* Agona, *S.* Anatum, *S.* Bovismorbificans, *S.* Braenderup, *S.* Brandenburg, *S.* Bredeney, *S.* Derby, *S.* Enteritidis, *S.* Hadar, *S.*Heidelberg, *S.* Infantis, *S.* Kentucky, *S.* Kottbus, *S.* Livingstone, *S.* London, *S.* Mbandaka, *S.* Montevideo, *S.* Newport, *S.* Rissen, *S.* Saintpaul, *S.* Typhimurium and *S.* Virchow. Albeit important in humans in most of the 24 countries, *S.* Dublin, *S.* Ohio and S. *S*tanley were excluded because *S.* Stanley was not isolated from the selected reservoirs, while *S.* Dublin

and *S*. Ohio became irrelevant after cattle was excluded. The building structure of the final *Salmonella* dataset (trade data not included) is shown in Figure 2.

DISCUSSION

This study presented the officially reported data available to be used in a EU-level source attribution model based on microbial subtyping [20]. Challenges associated with the use of those data were also presented, and solutions were proposed. The data available were retrieved from multiple sources and presented varied levels of quality and completeness. Although TESSy and EFSA collect and organize the data at EU-level in a harmonized way, the primary information is collected in different countries, which have their particular approaches and methods for data collection and management. Non-EU European countries such as Switzerland and Norway are also a source of data heterogeneity, as they participate in some studies and report partial data, for example, to the EUROSTAT production database, but not to the trade database. This variability made several data management steps necessary.

The variability observed in the number of reported human *Salmonella* infections reflects true differences in the burden of salmonellosis across countries, but also differences in foodborne disease surveillance systems in MSs and different levels of underreporting. The loss of data at various points along the surveillance chain from patient to official statistics is recognized in all countries [25], and multiplying factors [26] were used to try to compensate the occurrence of underreporting. Limitations and assumptions of the use of those factors should be discussed, as they were calculated based on Swedish cases [26], which came from a system where underreporting is also expected to occur. By using the infection rates in returning travelers to calculate incidences for the local population in the countries visited, it was assumed that the eating habits and other exposures of Swedish travelers are the same as the locals', also disregarding local levels of acquired immunity and differences in circulating strains. Similar considerations must be done regarding the use of a Dutch population-based *Salmonella* prevalence study as a reference to estimate the underreporting in the other countries, and a full discussion of the limitations can be found in Havelaar et al. [26]. Furthermore, this adjustment is expected to affect the relative importance attributed to the different sources by the model at EU level, as it affects the contribution of each country to the total burden of salmonellosis in the EU.

Information about travelling within or outside Europe was not available in a representative manner, and it was not possible estimate additional "extra" intra-EU travelers because the proportion of reported cases with missing travel information varied substantially, being 100% in some countries. Thus, it had to be assumed that all reported cases with missing travel information were domestically acquired, which is expected to be an overestimation, since travel information as reported to TESSy is often incomplete and may not reflect the true relation between travel and domestic cases [4].

Concerning the animal data, the panel of participating MSs varied with each BS, as countries have the right to refuse participation in the EU-wide Baseline Studies. The admittance of new MSs to the EU also generates different lists of reporting countries for each animal source, as data were collected in different years. The resulting data gaps were, when possible, filled with information from the EUSR. There are currently no EU-wide studies on the baseline prevalences of *Salmonella* in cattle and no harmonized monitoring in place, which is the main reason why this reservoir was excluded from this study. However, this is not expected to have a large impact on the model, as national attribution studies have suggested that the contribution from the cattle reservoir in general is low when compared to the other sources [16].

Data were also heterogeneous in regards to serotyping information and reporting of aggregated data or data with no or sparse serotyping information for both humans and animals. To deal with missing or aggregated information, countries were approached directly for more complete datasets, and records were reassigned based on the serovar distributions observed in available data or external reference datasets (e.g., WHO GFN/CDB). One limitation of this approach is that any emergence of new serovars or other profile fluctuations may be lost, particularly in situations where a whole year of typing is missing and the records are reassigned based on data from previous years. Therefore, serovar reassignment is considered a large source of uncertainty around the final data, and it is proposed that future models use a stochastic approach for reassigning, allowing for this uncertainty to be expressed and quantified.

The consumption dataset presented a special challenge, as it had to be often built based on estimates from surrogate trade data, and an evaluation of the quality of the trade data collected by EUROSTAT has revealed major and persistent inconsistencies in the various MSs intra-EU trade statistics [32]. However, the

comparison with the WHO GEMS/Food data showed that this approach produced valid results, as 19 out of 24 countries had a PSI of 0.9 or higher and three were larger than 0.8, suggesting that the consumption profiles composed using EUROSTAT data are highly similar to the original GEMS/Food profiles for most countries. An exception is noted for Cyprus, which is likely to be a reflection of the large proportion of data that needed extrapolation, and which may have an effect on the attribution outcomes for that country. Nonetheless, the dataset as a whole, there showed 91% similarity.

In conclusion, as long as a thorough data evaluation is performed and specific countries and reservoirs with insufficiently representative data are excluded, public surveillance and monitoring data from multiple countries can be used for scientific purposes, particularly for microbial subtyping-based source attribution methods. This could be a first step to the conduction of source attribution studies in countries where no country-wide baseline studies or serological surveys have been conducted, but where programs for *Salmonella* monitoring in food or surveillance in humans are currently up and running.

ACKNOWLEDGEMENTS

We would like to acknowledge the staff of EFSA's Task Force of Zoonoses Data Collection for providing the original datasets necessary to this study. The EU model was developed with funding from contract CT/EFSA/Zoonoses/2010/02 between EFSA and the DTU National Food Institute, in relation to Question n° EFSA-Q-2010-00685. We also acknowledge Timour Koupeev from Vose Risk Consulting for the collaboration on the management of the EUROSTAT data.

REFERENCES

1. Greger M. 2007. The human/animal interface: emergence and resurgence of zoonotic infectious diseases. Critical Reviews in Microbiology 2007; 33: 243-299.

2. World Health Organization. The world health report 2007. (http://www.who.int/whr/2007/en/index.html). Geneva, Switzerland: World Health Organization. Acessed 7 July 2011.

3. Tauxe RV, *et al*. Evolving public health approaches to the global challenge of foodborne infections. International Journal of Food Microbiology 2010; 139:S16–S28.

4. European Food Safety Authority. The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2009. EFSA Journal 2011; 9:2090.

5. European Food Safety Authority. Report on the analysis of the baseline study on the prevalence of *Salmonella* in holdings of laying hen flocks of *Gallus gallus*. EFSA Journal 2007; 97:1-84.

6. European Food Safety Authority. Report of the task force on zoonoses data collection on the analysis of the baseline survey on the prevalence of *Salmonella* in slaughter pigs in the EU, 2006-2007. Part A: *Salmonella* prevalence estimates. The EFSA Journal 2008; 135:1-111.

7. European Food Safety Authority. Report of the task force on zoonoses data collection on the analysis of the baseline survey on the prevalence of *Salmonella* in turkey flocks, in the EU, 2006-2007. Part A: *Salmonella* prevalence estimates..The EFSA Journal 2008; 134:1-224.

8. European Food Safety Authority. Report of the task force on zoonoses data collection on the analysis of the baseline survey on the prevalence of *Salmonella* in broiler flocks of *Gallus gallus* in the EU, 2005-2006. Part A: *Salmonella* prevalence estimates. The EFSA Journal 2007; 98:1-85.

9. European Food Safety Authority. EFSA Panel of Biological Hazards (BIOHAZ); Scientific Opinion on a request from European Commission on quantitative estimation of the impact of setting a new target for the reduction of *Salmonella* in breeding hens of *Gallus gallus*. The EFSA Journal 2009; 1036:1-68.

10. European Food Safety Authority. EFSA Panel of Biological Hazards (BIOHAZ); Scientific Opinion on a quantitative estimation of the public health impact of setting a new target for the reduction of *Salmonella* in broilers. EFSA Journal 2011; 9:2106.

11. European Food Safety Authority. EFSA Panel of Biological Hazards (BIOHAZ); Scientific Opinion on an estimation of the public health impact of setting a new target for the reduction of *Salmonella* in turkeys. EFSA Journal 2012; 10:2616.

12. Kuchenmüller T, *et al.* Estimating the global burden of foodborne diseases – a collaborative effort. Eurosurveillance 2009; 14:1-4.

13. Pires SM, *et al*. Attributing the Human Disease Burden of Foodborne Infections to Specific Sources. Foodborne Pathogens and Disease 2009; 6:417-424.

14. European Food Safety Authority. EFSA Panel on biological hazards (BIOHAZ); Scientific opinion on a request from EFSA on overview of methods for source attribution for human illness from food borne microbiological hazards. The EFSA Journal 2008; 764:1-43.

15. Hald T, *et al*. A Bayesian approach to quantify the contribution of animal-food sources to human salmonellosis. Risk Analysis 2004; 24: 255-269.

16. Pires SM, Hald T. Assessing the differences in public-health impact of *Salmonella* subtypes using a Bayesian microbial subtyping approach for source attribution. Foodborne Pathogens and Diseases 2010; 7:143-151.

17. Whalström H, *et al.* Source attribution of human *Salmonella* cases in Sweden. Epidemiology and Infection 2011; 139: 1246-1253.

18. Pires SM, Attributing human salmonellosis and campylobacteriosis to food, animal and environmental sources (thesis). Copenhagen, University of Copenhagen. 2009. 157p.

19. Pires SM, *et al*. Using outbreak data for source attribution of human salmonellosis and campylobacteriosis in Europe. Foodborne Pathogens and Disease 2010; 7:1351-1361.

20. de Knegt LV. A multi-country approach for attributing human salmonellosis to animla reservoirs: global perspectives and application of surveillance data from the European Union (thesis). Kongens Lyngby, Denmark: Technical University of Denmark, 2013, 109p.

21. European Centre for Disease Prevention and Control. 2007. Surveillance of communicable diseases in the European Union: A long-term strategy 2008–2013. (http://ecdc.europa.eu/en/aboutus/Key%20Documents/08-13_KD_Surveillance_of_CD.pdf). Solna, Sweden: European Centre for Disease Prevention and Control, 2007. Acessed 02 November 2012.

22. European Food Safety Authority. Analysis of the baseline survey on the prevalence of *Campylobacter* in broiler batches and of *Campylobacter* and *Salmonella* on broiler carcasses in the EU, 2008 - Part A: *Campylobacter* and *Salmonella* prevalence estimates. EFSA Journal 2010; 8:1503.

23. European Food Safety Authority. Analysis of the baseline survey on the prevalence of *Salmonella* in holdings with breeding pigs in the EU, 2008 - Part A: *Salmonella* prevalence estimates. EFSA Journal 2009; 7:1377.

24. Statistical Office of the European Union (EUROSTAT). (epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database). Acessed 25 March 2011.

25. Wheeler JG, *et al*. Study of infectious intestinal disease in England: rates in the community, presenting to general practice, and reported to national surveillance. British Medical Journal 1999; 318:1046-1050.

26. Havelaar AH, *et al*. Estimating the true incidence of campylobacteriosis and salmonellosis in the European Union, 2009. Epidemiology and Infection 2012; 141: 293-302.

27. World Health Organization Collaborating Center for Reference and Research on *Salmonella*. Antigenic formulae of the *Salmonella* serovars. 9th Edition. Paris, France: Institut Pasteur, 2007.

28. World Health Organization Global Foodborne Infections Network Country Databank (WHO/GFN CDB). (http://thor.dfvf.dk/gss) Acessed 15 November 2010.

29. European Food Safety Authority. The community summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in the European Union in 2007. EFSA Journal 2009; 223:312.

30. European Food Safety Authority. The community summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in the European Union in 2008. EFSA Journal 2010; 8:1496.

31. Association of Poultry Processors and Poultry Trade in the EU Countries. Annual Report 2009. Brussels, Belgium: Association of Poultry Processors and Poultry Trade in the EU Countries.

32. European Food Safety Authority. Collection and routine analysis of import surveillance data with a view to identification of emerging risks. EFSA Journal 2010; 8:1531.

33. World Health Organization Global Environment Monitoring System Food Consumption Cluster Diets. GEMS/Food regional diets: regional per capita consumption of raw and semi-processed agricultural commodities / prepared by the Global Environment Monitoring System/Food Contamination Monitoring and Assessment Programme (GEMS/Food). Geneva, Switzerland: World Health Organization 2006.

34. Rosef O, *et al*. Serotyping of *Campylobacter jejuni*, *Campylobacter coli* and *Campylobacter laridis* from domestic and wild animals. Applied Environmental Microbiology 1985; 49:1507–1510.

35. Feinsinger P, Spears EE, Poole RW. A simple measure of niche breadth. Ecology 1981; 62:27–32.

36. Mullner P, *et al*. Source attribution of food-borne zoonoses in New Zealand: a modified Hald model. Risk Analysis 2009; 29: 970-984.

37. Mullner P, *et al*. Molecular epidemiology of *Campylobacter jejuni* in a geographically isolated country with a uniquely structured poultry industry. Applied Environmental Microbiology 2010; 76:2145-2154.

38. European Food Safety Authority. EFSA Panel on biological hazards (BIOHAZ); Scientific opinion on monitoring and assessment of the public health risk of "*Salmonella* Typhimurium-like" strains. EFSA Journal 2010; 8:1826.

39. European Food Safety Authority. The community summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in the European Union in 2006. EFSA Journal 2007; 130:352.

40. David J, Attribution des cas de salmonelloses humaines aux différentes filières de production animale en France: adaptabilité et robustesse du modèle bayésien d'attribution par typage microbiologique (thesis). Vannes, France: Université Européenne de Bretagne, 2009, 217p.

Tables

Table 1. Availability of data from the different datasets by country.

Source	Data source[(a)]	Countries	Additional data sources
Laying hens	EUSR data 2008	AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GR, HU, IE, IT, LT, LU, LV, NL, NO, PL, PT, RO, SE, SI, SK, UK	
Cattle	EUSR data 2007-2009	AT, BE, BG, CH, CZ, DE, DK, EE, ES, FI, FR, GR, HU, IE, IT, LT, LU, LV, NL, NO, PL, PT, RO, SE, SI, SK, UK	FR: David, J (2009); LV: EUSR 2006
Pigs	BS 2006, lymph node	AT, BE, BG, CY, CZ, DE, DK, EE, ES, FI, FR, GR, HU, IE, IT, LT, LU, LV, NL, NO, PL, PT, SE, SI, SK, UK	
Broiler	BS 2008, carcasses	AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GR, HU, IE, IT, LT, LU, LV, MT, NL, NO, PL, PT, RO, SE, SI, SK, UK	GR: BS 2005/6
Turkey	BS 2006, Fattening turkeys	AT, BE, BG, CY, CZ, DE, DK, EE, ES, FI, FR, GR, HU, IE, IT, LT, LU, LV, NL, NO, PL, PT, SE, SI, SK, UK	EE: EUSR 2006; LU: EUSR 2008 LV: EUSR 2006;
Human cases	Foodborne outbreak data, 2007-2009	AT, BE, CH, CZ, DE, DK, EE, ES, FI, FR, HU, IE, LT, LV, NL, NO, PL, PT, RO, SE, SI, SK	
	TESSy case-based and aggregated data, 2007-2009[(b)]	AT, BE, BG, CY, CZ, DE, DK, EE, ES, FI, FR, GR, HU, IE, IT, LT, LU, LV, MT, NL, NO, PL, PT, RO, SE, SI, SK, UK	
	National monitoring and laboratory surveillance data 2007-2009[(c)]	PL, PT, NL, IT, DE	

(a) If data were missing from a specific source in a country, used surrogate data sources are indicated.

(b) Bulgaria reported human cases, but no serovar information was available.

(c) Obtained through direct contact with Member States.

Table 2. Number of countries in which each serovar is present by data source.

Serovar	Number of countries [a]							Total sources
	TESSy	FBO[b]	Broilers	Pork	Turkey	Layers	Cattle	(n=7)
S. Enteritidis*	27	19	15	18	17	20	25	7
S. Typhimurium*	27	16	12	23	12	16	25	7
S. Virchow*	26	3	7	5	2	11	1	7
S. Infantis*	26	3	15	16	4	13	8	7
S. Newport*	26	4	3	7	9	7	2	7
S. Derby*	25	1	3	19	11	5	10	7
S. Agona*	24	2	10	12	8	9	5	7
S. Hadar*	24	2	10	3	10	7	3	7
S. Bredeney	24	2	8	9	6	5	2	7
S. Kentucky*	22	0	6	0	1	2	0	4
S. Braenderup	23	0	2	3	2	7	1	6
S. Saintpaul	22	1	2	2	11	4	1	7
S. Brandenburg	23	1	1	7	0	4	2	6
S. Montevideo	22	0	8	9	3	10	4	6
S. London	22	0	1	8	1	1	2	6
S. Bovismorbificans*	22	4	0	7	0	1	2	5
S. Stanley	21	1	0	0	0	0	0	2
S. Mbandaka	20	1	11	4	2	10	4	6
S. Rissen	20	0	0	5	0	7	5	4
S. Anatum	19	1	5	9	4	5	1	7
S. Livingstone	19	0	5	5	1	10	4	6
S. Heidelberg	20	2	2	1	3	3	0	6
S. Ohio	18	1	5	6	0	3	2	6
S. Kottbus	18	0	4	2	9	1	2	6
S. Dublin	16	1	0	2	0	1	14	5

(a) n(TESSy)=27; n(FBO)=22; n(Broilers)=29; n(Pork)=26; n(Turkey)=26; n(Layers)=28; n(Cattle)=27.

(b) FBO: Foodborne outbreaks.

Table 3. Number and percentage of reassigned records in humans.

Country	Incomplete identification						Aggregated data[(d)]		Unknown[(e)]		Total		
	Species/genus[(a)]		Subspecies[(b)]		Serogroup[(c)]						Reported	Reassigned	
	n	%	n	%	n	%	n	%	n	%		n	%
AT			2	0.02	132	1.56	287	3.38	362	4.27	8,487	783	9.23
BE							172	1.55			11,066	172	1.55
BG	-	-	-	-	-	-	-	-	-	-	3,899	-	-
CY	2	0.42			9	1.91			101	21.44	471	112	23.78
CZ									586	1.51	38,842	586	1.51
DE			462	0.36	8,057	6.33	5,782	4.54	1,628	1.28	127,330	15,929	12.51
DK			2	0.03	3	0.04	25	0.33	342	4.56	7,497	372	4.96
EE					25	1.86	28	2.09			1,341	53	3.95
ES							2,504	20.81	2,091	17.38	12,033	4,595	38.19
FI	19	0.23	3	0.04	23	0.28	6	0.07	22	0.27	8,228	73	0.89
FR							2,185	10.75			20,319	2,185	10.75
GR					104	5.40	3	0.16	1,309	67.93	1,927	1,416	73.48
HU			57	0.30	191	1.00	908	4.76	2	0.01	19,091	1,158	6.07
IE	1	0.08					11	0.87	68	5.38	1,264	83	6.57
IT	25	0.24			6	0.06			1,080	10.58	10,205	1,111	10.89
LT					56	0.73	156	2.04	191	2.50	7,643	403	5.27
LU									63	13.15	479	63	13.15
LV							53	1.99	608	22.81	2,665	661	24.80
MT	20	5.39							40	10.78	371	60	16.17
NL			210	5.04			84	2.02			4,168	294	7.05
PL							1204	3.89			30,963	1,204	3.89
PT											1,513	0	0.00
RO							1,218	51.81	766	32.58	2,351	1,984	84.39
SE			68	0.60			411	3.65	307	2.73	11,265	786	6.98
SI					63	2.10					3,002	63	2.10
SK	3	0.02			154	0.79	84	0.43	87	0.45	19,399	328	1.69
UK	7	0.02			149	0.41	4	0.01	1,009	2.75	36,666	1,169	3.19
EU total	77	0.02	804	0.20	8,975	2.29	15,125	3.85	10,662	2.72	392,485	35,643	9.08
CH	-	-	-	-	-	-	-	-	-	-	-	-	-
NO							21	0.44	10	0.21	4825	31	0.64
Total	77	0.02	804	0.20	8,975	2.26	15,146	3.81	10,672	2.69	397,310	35,674	8.98

(a) *Salmonella* spp, *Salmonella enterica*, *Salmonella* not typed, *Salmonella* untyped

(b) *Salmonella enterica enterica*, Subspecies I

(c) B, C, D, E, D1, C1, C2-C3, D1, E1

(d) ”Others”, ”Other serovars”

(e) ”Unknown”

Table 4. Number and percentage of reassigned records in foodborne *Salmonella* outbreaks.

Country	Reported	Incomplete identification				Total		
		Species/genus[(a)]		Serogroup[(b)]		Reported	Reassigned	
		n	%	n	%		n	%
AT	Yes					421	0	0.00
BE	Yes					91	0	0.00
BG	No					-	-	-
CY	No					0	0	0.00
CZ	Yes					337	0	0.00
DE	Yes	13	0.55			2,383	13	0.55
DK	Yes					2,224	0	0.00
EE	Yes					157	0	0.00
ES	Yes					469	0	0.00
FI	Yes					189	0	0.00
FR	Yes	1218	46.68			2,609	1,218	46.68
GR	No					0	0	0.00
HU	Yes	86	4.48			1,921	86	4.48
IE	Yes					67	0	0.00
IT	No					0	0	0.00
LT	Yes					371	0	0.00
LU	No					0	0	0.00
LV	Yes	201	39.26			512	201	39.26
MT	No					0	0	0.00
NL	Yes	12	1.71	26	3.71	700	38	5.43
PL	Yes			29	0.55	5,310	29	0.55
PT	Yes					90	0	0.00
RO	Yes	26	5.95			437	26	5.95
SE	Yes	8	2.94			272	8	2.94
SI	Yes					692	0	0.00
SK	Yes					583	0	0.00
UK	No					0	0	0.00
EU total	-	1,564	7.89	55	0.28	19,835	1,619	8.16
CH	Yes					6	0	0.00
NO	Yes					95	0	0.00
Total	-	1,564	7.85	55	0.28	19,936	1,619	8.12

(a) *Salmonella enterica enterica*, Subspecies I

(b) B, C, D, E, D1, C1, C2-C3, D1, E1

Table 5. Number of cases reported in the original datasets as travel-related, domestic or unknown and the total used in the model, assuming that any case not specifically mentioned as travel-related was domestic.

Country	Reported			Total used	
	Travel	Domestic	Unknown	Travel	Domestic
AT	988	7,499	0	988	7,499
BE	0	11,066	0	0	11,066
BG	-	-	-	-	-
CY	18	428	25	18	453
CZ	657	38,185	0	657	38,185
DE	6,683	114,362	6,285	6,683	120,647
DK	1,366	2645	3,486	1,366	6,131
EE	95	1246	0	95	1,246
ES	0	12,033	0	0	12,033
FI	6,845	1059	324	6,845	1,383
FR	0	0	20,319	0	20,319
GR	45	1763	119	45	1,882
HU	29	19,062	0	29	19,062
IE	384	343	537	384	880
IT	132	692	9,381	132	10,073
LT	21	0	7,622	21	7,622
LU	46	431	2	46	433
LV	32	1,817	816	32	2,633
MT	4	365	2	4	367
NL	497	3,671	0	497	3,671
PL	16	0	30,947	16	30,947
PT	5	0	1,508	5	1,508
RO	0	0	2,351	0	2,351
SE	8,752	2,207	306	8,752	2,513
SI	0	0	3,002	0	3,002
SK	146	19,253	0	146	19,253
UK	8,921	8,084	19,661	8,921	27,745
EU total	35,682	246,211	106,693	35,682	356,803
CH	-	-	-	-	-
NO	3,721	871	233	3,721	1,104
Total	39,403	247,082	106,926	39,403	357,907

Table 6: Number of sampling units submitted and positivity percentages in animal reservoirs in the EU and Norway.

Country	Broiler carcasses[(a)]			Pigs – lymph nodes			Laying hen flocks			Turkeys – fattening flocks			Cattle [(b)]		
	Submitted	Positives		Submitted	Positives		Submitted	Positives		Submitted	Positives		Submitted	Positives	
		n	%		n	%		N	%		N	%		N	%
AT	408	10	2.5	617	13	2.1	1,966	49	2.5	1,010	141	14.0	3,037	12	0.4
BE	380	77	20.3	601	78	13.0	649	76	11.7	370	40	10.8	81	81	100.0
BG	316	85	26.9	176	35	19.9	119	0	0.0	85	0	0.0	477	3	0.6
CY	357	38	10.7	359	47	13.1	40	5	12.5	70	28	40.0	-	-	-
CZ	422	23	5.5	654	38	5.8	449	40	8.9	970	192	19.8	696	24	3.4
DE	432	76	17.6	2,567	325	12.7	6304	220	3.5	1,475	108	7.3	4,053	163	4.0
DK	396	0	0.0	998	80	8.0	508	3	0.6	294	1	0.3	7,915	9	0.1
EE	102	0	0.0	420	27	6.4	52	4	7.7	2	0	0.0	1,550	10	0.6
ES	389	58	14.9	2,621	806	30.7	845	376	44.5	1,910	747	39.1	258	29	11.2
FI	369	0	0.0	419	0	0.0	950	1	0.1	675	0	0.0	3,415	7	0.2
FR	422	32	7.6	1,163	215	18.5	3067	187	6.1	1,630	157	9.6	-	-	2.4
GR	1,215	180	14.8	345	73	21.2	112	35	31.3	220	16	7.3	56	1	1.8
HU	321	275	85.7	656	75	11.6	866	101	11.7	1,465	915	62.5	178	31	17.4
IE	394	39	9.9	422	65	15.4	204	2	0.98	1,295	294	22.7	10,121	430	4.2
IT	393	66	16.8	709	116	16.4	821	171	20.8	1,370	277	20.2	1,797	17	0.9
LT	374	26	6.9	461	8	1.7	13	0	0.0	315	14	4.4	172	2	1.2
LU	13	0	0.0	313	50	16.0	7	1	14.3	1	0	0.0	83	7	8.4
LV	122	6	4.9	392	21	5.4	69	14	20.3	1	0	0.0	25	0	0.0
MT	367	77	21,0	-	-	-	-	-	-	-	-	-	-	-	-
NL	429	43	10.0	1,087	92	8.5	2346	62	2.6	860	77	9.0	330	18	5.5
PL	419	107	25.5	1,176	75	6.4	1533	192	12.5	1,610	285	17.7	130	0	0.0
PT	421	47	11.2	658	156	23.7	227	83	36.56	525	26	5.0	56	0	0.0
RO	357	17	4.8	-	-	-	-	-	-	-	-	-	521	3	0.6
SE	410	1	0.2	394	6	1.5	724	5	0.7	70	0	0.0	3,728	60	1.6
SI	413	7	1.7	431	27	6.3	172	18	10.5	655	100	15.3	386	1	0.3
SK	422	91	21.6	385	30	7.8	138	10	7.2	125	15	12.0	95	0	0.0
UK	401	14	3.5	639	139	21.8	5523	67	1.2	1,570	401	25.5	895	895	100.0
EU Total	9,249	1,215	13.1	18,663	2,596	13.9	27,704	1630	5.9	18,514	3,834	20.7	40,055	1,803	4.5
NO	396	0	0.0	408	1	0.2	1080	0	0.0	360	0	0.0	2,589	1	0.0
Total	10,035	1,225	12.2	19,072	2,598	13.6	28,784	1630	5.7	18,849	3,834	20.3	42,644	1,804	4.2

(a) In the specific case of Greece, broiler flocks. (b) In the specific case of Denmark, carcass samples collected at the slaughterhouse.

Table 7. Number and percentage of records reassigned to serovars in animal reservoirs.

	Country	Incomplete identification						Aggregated [(d)]		Total		
		Species/genus[(a)]		Subspecies[(b)]		Serogroup[(c)]				Positives	Reassigned	
		n	%	n	%	n	%	n	%		n	%
Broilers	BE	15	19.48							77	15	19.48
	IT	13	19.70							66	13	19.70
	LT	15	57.69							26	15	57.69
	MT	10	12.99							77	10	12.99
	NL	1	2.33							43	1	2.33
Pigs	BG			4	11.43					35	4	11.43
	CY	5	10.64	3	6.38	1	2.13			47	9	19.15
	DE	5	1.54			64	19.69			325	69	21.23
	EE			4	14.81					27	4	14.81
	ES	62	7.69							806	62	7.69
	FR	5	2.33							215	5	2.33
	GR	3	4.11	8	10.96					73	11	15.07
	IE	1	1.54							65	1	1.54
	IT	41	35.34	6	5.17					116	47	40.52
	LV	2	9.52							21	2	9.52
	NL	2	2.17	2	2.17					92	4	4.35
	SI	4	14.81							27	4	14.81
Turkeys	CY					5	17.86			28	5	17.86
	DE					11	10.19			108	11	10.19
	DK	1	100.00							1	1	100.00
	HU	1	0.11	2	0.22					915	3	0.33
	IT			8	2.89					277	8	2.89
	SI					1	1.00			100	1	1.00
Layers	AT	2	4.08							49	2	4.08
	BE	3	3.95			3	3.95			76	6	7.89
	CY					1	20.00			5	1	20.00
	DE	13	5.91					23	10.45	220	36	16.36
	ES	186	49.47							376	186	49.47
	FR	20	10.70					6	3.21	187	26	13.90
	HU							26	25.74	101	26	25.74
	IT							115	67.25	171	115	67.25
	PL							29	15.10	192	29	15.10
	PT							9	10.84	83	9	10.84
	UK							16	23.88	67	16	23.88
Bovines	BE	3	3.70			4	4.94			81	7	8.64
	DE	4	2.45					36	22.09	163	40	24.54
	DK	4	44.44							9	4	44.44
	ES	13	44.83							29	13	44.83
	HU	25	80.65							31	25	80.65
	IT	4	23.53							17	4	23.53
	LU	1	14.29							7	1	14.29
	NL	1	5.56							18	1	5.56
	SE	6	10.00							60	6	10.00
	UK	824	92.07							895	824	92.07

(a) *Salmonella* spp, *Salmonella enterica*, *Salmonella* not typed, *Salmonella* untyped
(b) *Salmonella enterica enterica*, Subspecies I
(c) B, C, D, E, D1, C1, C2-C3, D1, E1
(d) ”Others”, ”Other serovars”

Table 8. Comparison of the relative proportion of consumption of pork, poultry meat and table eggs in the WHO GEMS/Food data and the surrogate values calculated from EUROSTAT data.

Country	WHO GEMS/Food (%)			EUROSTAT (%)			PSI
	Poultry	Pig	Egg	Poultry	Pig	Egg	
AT	16,7	70,9	12,4	18,8	68,8	12,4	**0,98**
BE	32,3	50,5	17,2	28,7	58,1	13,2	**0,92**
CY	38,7	48,3	13,0	96,8	2,9	0,3	**0,42**
CZ	28,6	52,7	18,6	28,4	52,9	18,7	**1,00**
DE	17,4	67,0	15,6	24,1	63,2	12,7	**0,93**
DK	19,4	64,2	16,5	13,1	81,3	5,6	**0,83**
EE	33,5	47,6	18,8	33,4	49,7	16,9	**0,98**
ES	25,8	61,0	13,2	30,9	56,2	12,9	**0,95**
FI	25,8	58,7	15,5	24,5	49,9	25,6	**0,90**
FR	32,9	47,7	19,4	42,1	39,5	18,4	**0,91**
GR	31,5	53,1	15,4	33,2	47,9	18,9	**0,95**
HU	33,2	49,8	17,0	41,0	42,0	17,1	**0,92**
IE	36,3	54,7	9,0	40,9	45,7	13,4	**0,91**
IT	24,4	59,9	15,7	31,0	53,9	15,1	**0,93**
LT	24,6	51,4	23,9	30,7	51,1	18,2	**0,94**
LU	47,8	44,3	8,0	32,2	45,7	22,1	**0,84**
LV	30,3	44,7	25,0	33,6	43,0	23,4	**0,97**
NL	16,2	59,6	24,2	31,0	51,5	17,5	**0,85**
PL	23,8	61,7	14,5	31,3	56,6	12,0	**0,92**
PT	32,7	54,2	13,1	34,8	50,7	14,5	**0,97**
SE	20,9	61,3	17,8	22,3	58,6	19,1	**0,97**
SI	37,9	50,9	11,2	44,6	39,2	16,2	**0,88**
SK	36,5	45,8	17,7	28,2	48,7	23,1	**0,92**
UK	44,2	38,7	17,1	48,0	33,7	18,3	**0,95**
Overall PSI							**0.91**

Figures

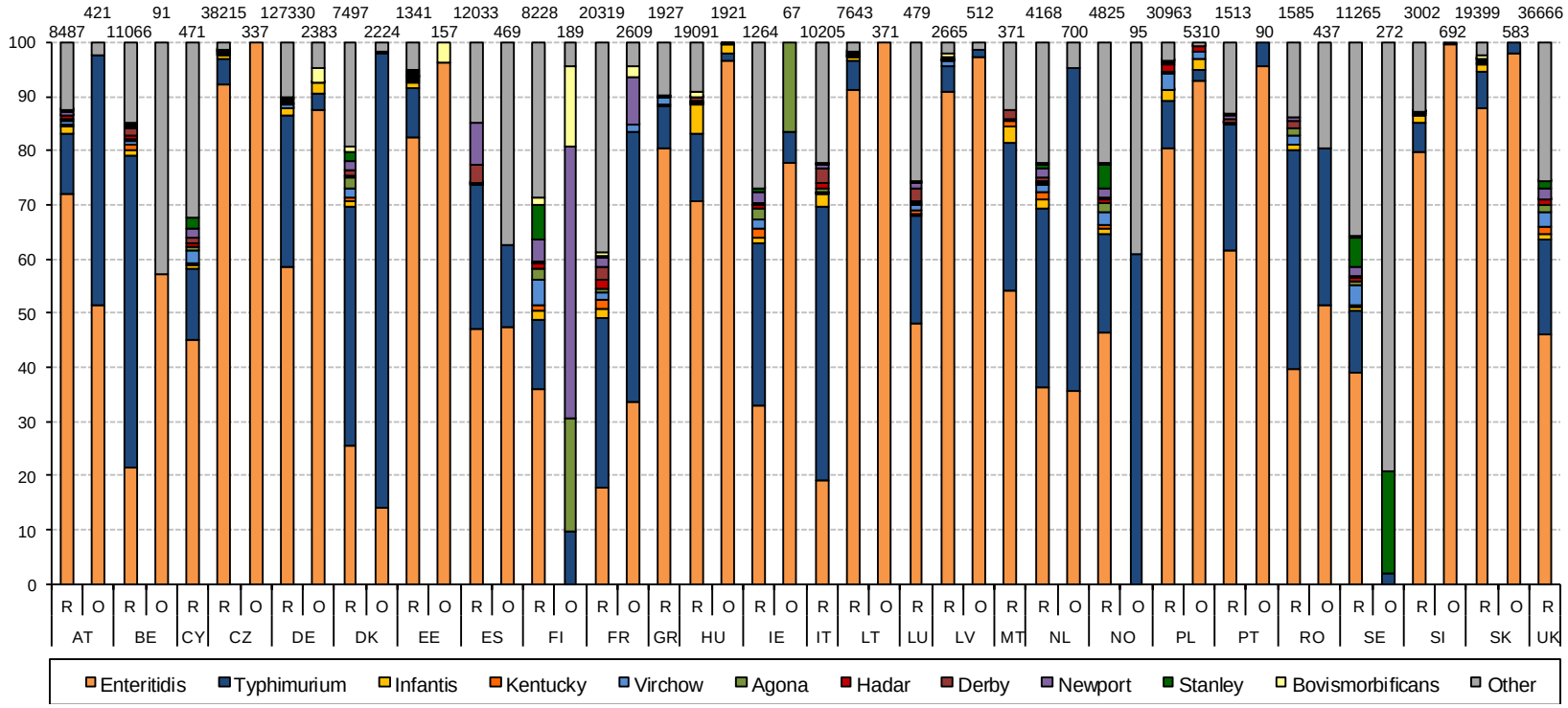


Figure 1: Relative proportions of the most frequent serovars in total reported (R) and outbreak (O) cases in humans in the EU and Norway, 2007-2009. The totals for each country in the datasets are shown at the top of the bar.

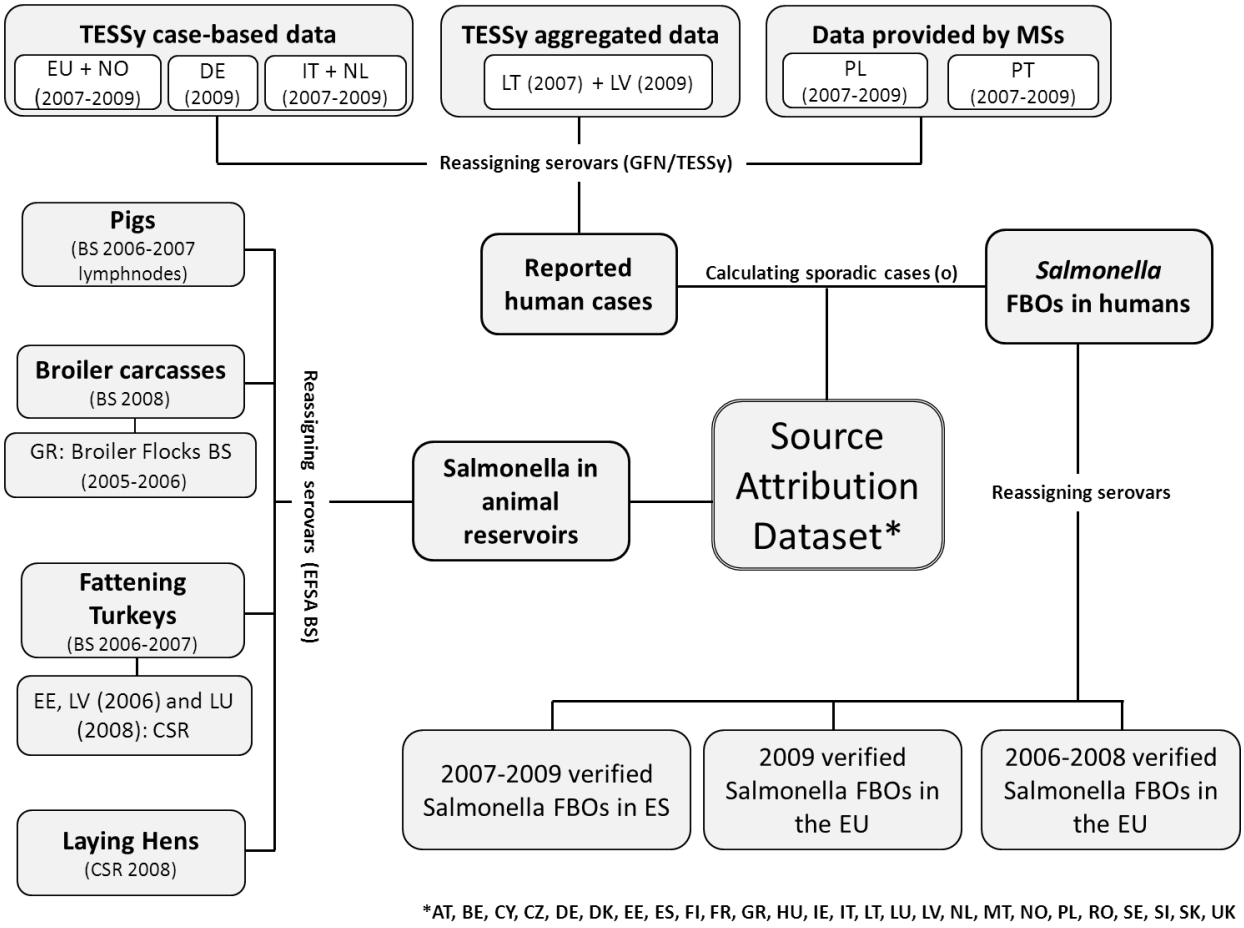


Figure 2. Diagram illustrating the construction of the final dataset for source attribution. For animal reservoirs and outbreaks, each gray block represents a dataset. For reported human cases, white blocks represent primary datasets originally provided to compose the gray blocks.

Manuscript II

Attributing foodborne salmonellosis in humans to animal reservoirs in the European Union using a multi-country stochastic model

L.V. DE KNEGT[(1)], S.M. PIRES[(1)], T. HALD[(1)]

[(1)],National Food Institute, Division of Epidemiology and Microbial Genomics

Technical University of Denmark

Author responsible for manuscript and reprint correspondence: L.V. DE KNEGT

Mailing address:

Technical University of Denmark, National Food Institute

Division of Epidemiology and Microbial Genomics

Mørkhøj Bygade 19, Building H, 2860 Søborg, Denmark

Email: ledkn@food.dtu.dk

Running head: *Salmonella* source attribution in the EU

Submitted to Epidemiology and Infection

Summary

A Bayesian modeling approach comparing the occurrence of *Salmonella* serovars in animals and humans was used to attribute salmonellosis cases to broilers, turkeys, pigs, laying hens, travel and outbreaks in 24 European Union countries. *Salmonella* data for animals and humans, covering the period from 2007 – 2009, were mainly obtained from studies and reports published by the European Food Safety Authority. Availability of food sources for consumption was derived from trade and production data from the European Statistical Office. Results showed layers as the most important reservoir of human salmonellosis in Europe, with 42.4% (7,903,000 cases, 95% Credibility Interval 4,181,000 – 14,510,000) of cases, 95.9% of which caused by *S.* Enteritidis. In Finland and Sweden, most cases were travel-related, while in most other countries the main sources were related to the laying hen or pig reservoir, highlighting differences in the epidemiology of *Salmonella*, surveillance focus and eating habits across the EU.

INTRODUCTION

Unsafe food is related to several kinds of diseases, ranging from diarrhoeal syndromes to various forms of cancer. It is estimated that food- or waterborne diarrhoeal diseases are responsible for 2.2 million deaths per year worldwide, 1.8 million of which are children [1]. *Salmonella enterica* is considered one of the leading causes of gastroenteritis and bacteremia in the world [2,3], being estimated to cause 93.8 million human cases and 155 thousand deaths every year [4]. In the European Union (EU), *S.* Enteritidis and *S.* Typhimurium are the most frequently reported serovars, but a wide range of others frequently cause disease in humans and thus are of public health significance [3,5]. Human infection is most often foodborne, but other routes of infection, namely contact with animals and environmental transmission, have been identified [6,7].

To design and prioritize effective food safety interventions, it is important to identify which foods are vehicles for specific illnesses [8]. This process is called *source attribution*, and it can be based on different approaches, such as analysis of outbreak data, analysis of sporadic cases, microbial subtyping, comparative exposure assessment, intervention studies and expert elicitations [8]. Methods for source attribution are intended to provide countries with tools for priority setting in relation to human foodborne and zoonotic diseases both at national and regional level, being a critical tool for decision-making aimed at reducing human zoonotic infections faster and more effectively [9].

Hald et al. [10] developed a Bayesian approach based on microbial subtyping for attribution of human cases of salmonellosis to animal reservoirs in Denmark. It made use of Denmark's extensive surveillance and data collection system to identify the main *Salmonella* subtypes responsible for human cases and compare them with the ones found in six animal-food sources. The model was further developed by Pires and Hald [11] to accommodate information from different time periods, and adapted by Mullner et al. [12] to apply it to *Campylobacter*.

Other EU Member States (MS) have performed *Salmonella* source attribution studies based on the cited methods, such as Sweden [13] and the Netherlands [14]. A EU-wide source attribution approach based on outbreak data was also developed [15]; this model attributed disease at the EU region level and did not

provide estimates at country level. So, the relative contribution of different food sources for human salmonellosis in the remaining individual countries within Europe had still not been assessed.

This paper presents a study in which the Hald model was adapted to use EU-harmonized data reported by 24 MSs to attribute human cases of salmonellosis to their respective animal reservoirs at country and EU-level.

METHODS

Data availability

All utilized data covered the period between 2007 and 2009. EU animal-food production and trade data were available as published by the Statistical Office of the European Union (EUROSTAT) [16]. Data on the prevalence of *Salmonella* serovars in animals and food were available from the EU-wide Baseline Studies (BS) conducted in different animal species [17, 18, 19, 20] and from the European Union Summary Reports (EUSR) as published by the European Food Safety Authority (EFSA) from 2006 to 2009 [21, 22, 23, 24]. Data on the number and serovar distribution of human cases reported to the European Surveillance System (TESSy) from 2007 to 2009 were extracted on 6th of July 2010 and provided by the European Centre for Disease Prevention and Control (ECDC) through EFSA, except for Poland and Portugal, which directly provided additional datasets with more detailed serovar information. Human data included both case-based and aggregated data and were complemented with other data sources (e.g. national monitoring or laboratory surveillance data not published in the EUSRs) when necessary and possible. One of the main obstacles for the use of these data is the underreporting of cases. It is generally understood that the real (and generally unknown) number of illnesses in the population is considerably larger than the number of cases reported in the surveillance system. Also, the level of underreporting varies strongly between countries, depending on differences in organization and effectiveness of local surveillance systems [25, 26]. This was taken into consideration by multiplying the country-specific underreporting factors (UFs) estimated by Havelaar et al. [27] to the reported sporadic cases. .The underreporting factors were fitted as lognormal distributions, following the methodology described in Hald et al. [28]. The number of cases originally reported in the datasets obtained, the underreporting factors and the resulting adjusted totals can be seen in Table 1.

Data management

Isolates not classified up to the serovar level or reported in aggregated form were reassigned to specific serovars according to proportions observed in previous studies, in the same dataset or in other references, depending on the availability of data in each case. Isolates classified as serogroups were distributed among serovars pertaining to them, in accordance with the Kauffman-White-Le Minor Scheme 9th edition [29]. For sporadic human cases, the main reference dataset used to obtain the proportions for the reassignment was the WHO Global Foodborne Infections Network (GFN) Country Databank (CDB) [30], which contains the 15 most commonly identified *Salmonella* serovars among human and non-human sources in 84 countries. Animal isolates were reassigned based on proportions found in the BS datasets. Isolates identified as monophasic variants of *S.* Typhimurium (e.g. *S.*1,4,[5],12:i:- or *S.*4,[5],12:i:-) were reassigned to *S.* Typhimurium [31]. Outbreak-related cases were reassigned using the proportions observed in the outbreak dataset, because some serovars may be more prone to generate outbreaks than others, and thus the proportions observed in reported sporadic cases may not apply. At the EU-level, a total of 9.1% of sporadic cases had to be reassigned to specific serovars, varying from zero in Portugal to 73.5% in Greece. Records with travel information referred as "unknown" and considered as domestic cases corresponded to 27% of all cases reported, varying from zero in Austria, Belgium, the Czech Republic, ,Estonia, Hungary, the Netherlands, Slovakia and Spain, to 100% in France. No outbreak cases were reported by Cyprus, Greece, Italy, Luxembourg, Malta or the United Kingdom. Among countries which reported outbreaks, the total percentage of reassigned cases was 8.2%, ranging from zero in 13 countries to 46.7% in France. Concerning the animal data, reassigned records corresponded to 4.4% of the total for broilers, 8.6% for pigs, 0.8% for turkeys, 27.8% for layers and 51.3% for cattle. The number of countries in which reassignments were necessary varied from five in broilers to 11 in pigs, and the largest reassigned percentage was observed for cattle in the UK (92.1%).

Concerning the consumption data, the domestic amount of a product available in a country was estimated as Domestic Production minus Export, whereas the amount of imported food available for consumption in MS A originating from MS B was estimated as Import minus Re-export (when re-export was relevant). That was

done in order to consider the intra-community food trade and its impact on the incidence of human salmonellosis in importing countries. Trade between EU countries and third countries was not considered. Based on data quality, food-animal sources included in the final model were broilers, pigs, turkeys and laying hens (as the animal reservoir for eggs). Since neither harmonized EU monitoring data nor BS data were available for the cattle reservoir, this source was excluded from the final model due to poor data quality, which would significantly compromise the validity of the model results. As for MSs, 24 were included in the model: Austria, Belgium, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden, The Netherlands and the United Kingdom. Twenty-two serovars were selected to be specifically addressed, based on their presence and importance in humans and in the main animal reservoirs in a five-year period: *S*. Agona, *S*. Anatum, *S*. Bovismorbificans, *S*. Braenderup, *S*. Brandenburg, *S*. Bredeney, *S*. Derby, *S*. Enteritidis, *S*. Hadar, *S*. Heidelberg, *S*. Infantis, *S*. Kentucky, *S*. Kottbus, *S*. Livingstone, *S*. London, *S*. Mbandaka, *S*. Montevideo, *S*. Newport, *S*. Rissen, *S*. Saintpaul, *S*. Typhimurium and *S*. Virchow. Albeit important in humans in most of the 24 countries, *S*.Dublin, *S*. Ohio and *S*. Stanley were not included in the list because *S*. Stanley was not isolated from the animal sources considered for the source attribution model, and *S*. Dublin and *S*. Ohio became irrelevant after the cattle reservoir was removed. Serovars not included in the above list were aggregated as "Others".

Data management was performed using SAS Enterprise Guide, SAS Institute, SAS/STAT® User's Guide, Version 8, Cary, NC: SAS Institute Inc., 1999. Data origin and countries providing information for each food-animal reservoir, reported human cases and cases related to foodborne *Salmonella* outbreaks are summarized in Figure 1.

Model overview

The presented approach for source attribution by microbial subtyping works by comparing the number of human cases caused by different subtypes of a pathogen with the distribution of the same subtypes in different food-animal sources, utilizing a collection of temporally and spatially related isolates from multiple sources and humans.

The model attributes sporadic domestic cases to food-animal sources. A sporadic case is defined as a subject that could not be associated with a recognized foodborne disease outbreak. Outbreak-related cases are added to the final results of the model, being attributed to the source implicated in the outbreak, if that is known. If not, they are considered outbreaks with unknown source. As *Salmonella* subtypes are clonally distributed among animal hosts [10], the model attribute cases at the animal reservoir level. This means that in general, cases caused by pork are attributed to pigs, eggs to layers, chicken meat to broilers and so on, but if a pork food preparation is contaminated during processing with a subtype originally found in broilers, the resulting cases are attributed to broilers, not pigs.

The model was built in a Bayesian framework based on the method described by Hald *et al.* [10]. In that model, *Salmonella* subtype distributions in animals in a given country in a certain time period are compared with the subtype distribution in humans in the same country in the same period.

The objective was to estimate the number of reported human cases that can be attributed to each source in each country, based on 1) the number of laboratory-confirmed infections caused by each *Salmonella* serovar in each country, including possible outbreak or travel information for each case, 2) prevalence of each serovar in the different sources in each country, and 3) amount of food source available for consumption in each country broken down by the country of origin. Due to the non-availability of animal data for the same years as the human data, it was decided to use a cross-sectional approach, using data from the EFSA baseline studies and assuming that the serovar profiles presented in them would be representative of the three-year period the human data referred to. The model was adapted to accommodate data from multiple countries, thereby adding a third dimension to the original model (in addition to subtype and food-animal source-related factors), and was based on the distribution of serovars in humans and food-animal sources. Another addition to the original model is the use of trade data as surrogate for consumption. This creates a scenario in which it is possible to differentiate the country of origin of the food from the country where the human cases were reported, and apply the corresponding country-specific *Salmonella* prevalences to the sources. As a consequence, it is also possible to estimate the number of cases reported in a country which are attributable to a source from other country(ies).

Model parameters and specifications

The model takes into account the number of cases caused by a serovar, the prevalence of each serovar in each source in each country, the underreporting multipliers in each country, and relative impact of a set of unknown factors, as described in Hald et al. [10]. The unknown factors were included as multi-parameter priors, and account for the differences in the ability of different subtypes to cause disease and of different sources to act as vehicles for infection. Multiple loops were included to accommodate data from the 24 countries. An overview of the model parameterization can be drawn as:

a_{cj} ~ Uniform (0,100)

q_i ~ Uniform (0,100)

λ_{ci} ~ Poisson (o_{ci}),

$\lambda_{ci} = \sum_{j=1\,k=1}^{n} \lambda_{ckji}$

$\lambda_{ckji} = p_{kij} * m_{ckj} * a_{cj} * q_i$

where: 1) λ_{ckji} is the expected number of cases per serovar *i* and source *j* reported in country *c* and caused by food produced in country *k*; 2) p_{kij} is the prevalence of serovar *i* in source *j* in country *k*; 3) m_{ckj} is the amount of source *j* available for consumption in country *c* produced in country *k*; when a source is domestically produced in the country of attribution, *c=k*; 4) a_{cj} is the source-dependent factor for source *j* in country *c*; 5) q_i is the subtype-dependent factor for serovar *i;* 6) and uf_c is the underreporting factor for the country of reporting. The source-dependent factor a_{cj} was assumed to vary between countries, accounting for variability in consumption patterns and preferences not captured by m_{ckj}, also including general variations between sources, *e.g*, bacterial load/concentration in the food and processing, handling or preparation practices. The subtype-dependent factor q_i is a one-dimensional parameter, meaning that it is a property of the *Salmonella* serovar and assumed independent of the country of infection. The q_i prior for *S.* Enteritidis is defined as 1, and all other q_i values are estimated relatively to this one. The amount of food source available for consumption in the country where a *Salmonella* case was reported considers both domestically produced and imported foods (m_{ckj}). The number of human sporadic and domestic cases attributed to each source per country (λc_{ji}) is estimated assuming a Poisson distribution of the observed number of sporadic cases per

subtype per country (o_{ci}). After attribution, sporadic reported cases were multiplied by the correspondent UF in each MS. Model parameters are presented in Table 2.

The model was built in WinBUGS 1.4 (http://www.mrc-bsu.cam.ac.uk/bugs/), which uses Markov Chain Monte Carlo (MCMC) with Gibbs sampling as a default to obtain summary values for posterior distributions. Five independent chains ran for 40,000 iterations each to obtain the values for a_{cj} and q_i. Each chain had a different set of starting values for the priors, widely dispersed in the target distribution. Chain convergence was monitored using the methods described by Gelman and Rubin [32] and was considered to have occurred when the variance between the different chains was no larger than the variance within each individual chain, and when the chains had reached a stable level.

RESULTS

The most important source of human salmonellosis at the EU level was estimated to be the laying hen reservoir (i.e. eggs), with 42.4% (7,903,000 cases, 95% Credibility Interval (CI) 4,181,000 – 14,510,000) of cases, followed by 31.1% attributed to pigs (5,800,000 cases, 95% CI 2,973,000 – 11,100,000). Broilers and turkeys were estimated to be less important sources of *Salmonella*, contributing with 12.6% (2,350,000 cases, 95% CI 736,300 – 6,194,000) and 3.8% (702,400 cases, 95% CI 325,500 – 1,590,000), respectively. A total of 1.6% (292,400 cases, 95% CI 150,700 – 562,700) of all salmonellosis cases were reported as being travel-related, and 0.1% (13,848) of cases were reported as being part of outbreaks with unknown source. Cases which could not be attributed to any of the sources included in the model corresponded to 8.5% of the total (1,578,000 cases, 95% CI 828,400 – 2,951,000).

The most important serovars contributing to human salmonellosis originating from the animal reservoirs are presented in Table 3. Of all *S.* Enteritidis infections, 63% (7,504,000 cases, 95% CI 3,964,000-13,770,000) were attributed to laying hens, whereas 90.8% of *S.* Typhimurium originated from pigs (2,950,000 cases, 95% CI 1,510,000-5,663,000). Compared to infections attributed to layers and pigs, a large proportion of cases were caused by other serovars in other sources, such as 4.5% *S.* Infantis in broilers (106,600 cases, 95% CI 32,560-284,500) and 9.2% *S.* Newport (226,296 cases, 95% CI 84,379-567,930) or 4.5% *S.*

Saintpaul (33,580 cases, 95% CI 18,052-62,443) in turkeys. In those sources, these serovars were not the most frequently associated with cases, but still constituted a significant burden.

When looking at attribution within specific countries, 13 MSs (Austria, Czech Republic, Estonia, Germany, Greece, Hungary, Latvia, Lithuania, Luxembourg, Slovenia, Slovakia, Spain and the United Kingdom) had the laying hen reservoir estimated as the most important source of salmonellosis. Pigs were the larger contributor for salmonellosis in eight (Belgium, Cyprus, Finland, France, Ireland, Italy, Poland and Sweden) MS, and the proportion of disease attributed to layers and pigs were similar in the Netherlands. Turkeys and broilers had a localized importance in Denmark and Portugal, respectively. The majority of *Salmonella* infections in Finland, Sweden and, to a lower extent, Denmark Ireland and the UK were reported as travel-related (Figure 2). Appendix A contains the country-specific attribution tables.

As mentioned earlier, a feature of this model is the ability to estimate the country of origin of cases attributed in other countries, as country-specific prevalences and amounts are used. When considering all sources together, Poland was estimated to be the most important source-country for human salmonellosis in the EU, contributing with 21.3% of cases (3,563,710 cases, 95% CI 911,750 – 10,818,900), followed by 18.4 from Spain (3,081,090 cases, 95% CI 898,170 – 9,056,800) and 14.5 from Portugal (2,422,142 cases, 95% CI 361,368 – 8,508,397) (Figure 16). Country-specific estimates with 95% Credibility Intervals are shown in Appendix B. Cases reported in the country of origin are also included in the total, which means that the 3,563,710 cases "originated" from Poland include cases reported in Poland, not only in other countries.

Looking at the numbers in Appendix B it can be seen that the impact of the country of origin varied with the source. As an example, 55.6% of cases (1,305,000 cases, 95% CI 198,500 – 4,535,000) attributed to broilers were estimated to originate" from Portugal, while cases attributed to turkeys were mostly related to Spain (43.1% or 302,600 cases, 95% CI 55,350 – 1,029,000) and pigs to Poland (24.2% or 1,402,000 cases, 95% CI 257,000 – 4,721,000) and Spain (22.5% or 1,306,000 cases, 95% CI 423,700 – 3,556,000). The majority of cases attributed to layers originated from Greece (21.5% or 1,701,000 cases, 95% CI 256,400 – 5,944,000), Spain (17.9% or 1,414,000 cases, 95% CI 406,000 – 4,286,000) and Poland (16.3% or 1,287,000 cases, 95% CI 492,000 – 3,162,000).

Concerning the factors simulated to estimate the ability of food sources to act as a vehicle for disease (a_{cj}) or of different serovars to cause disease (q_i), layers had the highest value of a_{cj} in 11 countries (Austria, Czech Republic, Estonia, Germany, Greece, Hungary, Lithuania, Luxembourg, Latvia, Slovenia and Slovakia) and turkeys in 10 (Belgium, Cyprus, Denmark, Finland, France, Ireland, the Netherlands, Spain, Sweden and the UK). In Italy and Poland, the highest a_{cj} was estimated for pigs, whereas in Portugal this happened for broilers. The highest values of q_i were estimated for *S.* Kentucky, *S.* Newport, *S.* Virchow and *S.* Typhimurium. Values estimated for a_{cj} and q_i are shown in Appendices C and D.

DISCUSSION

This study represents the first attempt to conduct source attribution of human salmonellosis in most European countries. Results suggest that layers were the most important source of salmonellosis in the EU in the study period, being responsible for over 40% of all *Salmonella* infections. At country level, it was estimated as the most important source in 13 out of 24 countries, followed by pigs, which was the most important source in eight countries. Turkeys were revealed as particularly important only in Denmark and broilers in Portugal. The identification of the most important sources of salmonellosis is a step for prioritization of actions and interventions aimed at reducing the public health burden of disease.

These attribution estimates took into account the amount of food produced and traded between countries as reported to the EUROSTAT database. The underlying assumption was that these data reflected the real flow of foodstuffs and consequent exposure in the countries. However, the dataset used was built based on production, imports, exports and poultry trade datasets, and their quality and consistency depend on factors as the recording and reporting of the information by the countries. It is an important feature in this model that the relative contribution of food-animals produced in different countries is dependent not only on the *Salmonella* prevalence in a source in an exporting country, but also on the amount imported from that country. This is a point in which the EU model differs from the way single-country models work: in a single-country model, m_j works as a subset of a_j, as they have the same dimensions (Hald, 2004; Pires and Hald, 2010; Whälstrom, 2011); for each source, there is only one value of *m* and one value for the prevalence of a subtype in that source. The m_j, therefore, has the role of weighting the contribution of the different sources,

which is, up to a point, already reflected in a_j, In the multi-country model, *m* in a reporting country is composed by subsets of *m* from different countries or origin of the food sources, each one with its own prevalence. For that reason, even if an exporting country has a very high prevalence in a source, this prevalence will have little impact in an importing country if the amount imported is very small, particularly if another country with a low prevalence exports very large amounts which can, ultimately, "dilute" the high prevalence found in the first country. In short, the amount imported ultimately drives the *m*p* in the model formula, particularly when large differences in trade volume are observed, and so the quality of the trade data could have a large impact on the observed results.

Travel-related cases had a localized importance in Northern Europe, notably in Scandinavian countries. Although data quality issues underline any interpretations of the travel data, these results are corroborated by other studies for at least two countries. A previous source attribution study in Sweden allocated 82% of *Salmonella* infections as travel-related [13], and results of the Danish source account for the same period [33] found a proportion of travel-related *Salmonella* cases varying between 22 and 46%, which, although higher than estimated by the EU model, accounted for the probability of a case with unknown travel information having been travelling abroad before onset of symptoms, and so add more "possible" travelers. Other countries, such as Spain, had zero cases attributed to international travel, as no travel information was reported. For this model, cases that were reported as acquired outside the country were considered as travel-related cases, and all cases without specific information otherwise were assumed to be domestically acquired. That resulted in the data available being dependent on the patients being asked whether they had been travelling abroad before onset of symptoms, and the information being registered centrally. For that reason, travel-related disease is expected to be underestimated. Differences between patients traveling within or outside Europe were not assessed, as this information was only available for few MS.

The use of underreporting factors has proved important when considering the effect of source and country contributions at EU level. This is particularly clear for broilers: this reservoir was the most important only in Portugal, but the use of an underreporting factor multiplied its impact within the EU by 2082.9, increasing both the relative contribution of broilers and of Portugal to the total cases of salmonellosis, when compared to the original numbers. A similar effect can be observed for the contribution of Greece to the total cases

attributed to layers. It should therefore be noted that most of the cases "originated" by countries with large underreporting factors were reported in those same countries, so one should be careful when interpreting these results as countries "exporting" cases to the rest of the EU.

As there was a large variation in the availability of data from the EFSA BS or EU- harmonized monitoring and surveillance of food sources between MSs, only broilers, laying hens, pigs and turkeys could be included in the model. This can result in the misplacing of some cases when their "right" source is not included As an example, it is expected that some cases that should be attributed to beef could be attributed to pigs instead, as *S*. Typhimurium is a common serovar in both sources. It should be noted, though, that when the Danish model started being applied, it only included five sources, and it was still a powerful tool in guiding the decisions for the targeted actions regarding broilers, pigs and table eggs that dramatically decreased the prevalence of *Salmonella* in these sources in the last decade [34, 35]. Fruits and vegetables, which are also recognized as sources of salmonellosis, were not included. This happened because the approach employed attributes cases to the original animal reservoirs, meaning that infections caused by fruits and vegetables contaminated with faeces from production animals would be traced to the animal reservoir.

The use of serovar as subtyping level, which resulted from the scarcity or absence of data on further subtyping levels (phage typing, antimicrobial resistance profiles), can also result in mis-attribution of cases. A good example is *S*. Enteritidis, which is present in all sources [17, 18, 19, 20]. Without more specific differentiation between subtypes found in each reservoir, cases are likely to be "cross-attributed" among sources. In countries where travel information was not provided, the mis-attribution of *S*. Enteritidis cases may include the attribution of cases which are actually travel-related to the animal reservoirs. In MSs with reasonably good travel data it can be seen that a large proportion of the *S*. Enteritidis infections are linked to travel, indicating that the same situation could be found in the MSs with poor or no travel data. In that scenario, travel-related cases would be wrongly attributed to one of the sources included in the model, as also observed by Hald et al. [28].

Despite data limitations and the consequent uncertainty in the results, the source attribution estimates are considered valid as a first indication of which sources are most important for human salmonellosis in several countries. Limitations include the variability in the human surveillance systems in place in the countries, as

well as the different details with which serovar information is reported for both human and animal-food sources. Such uncertainties cannot be statistically quantified, but should be kept in mind when interpreting the results. The relative importance of different food-animal sources was found to vary between countries according to differences in prevalences, trade and consumption patterns and preferences, as well as animal and food production systems, also highlighting regional differences in the focus of surveillance systems in place in EU Member States. The results are expected to be useful for the delineation of risk management strategies in the EU, and if the model is applied on a regular basis in upcoming years, it would be possible to analyse results over the years, for example, to evaluate the impact of implemented control, which would also be a way of validating the results.

ACKNOWLEDGEMENTS

We would like to acknowledge the staff of EFSA's Task Force of Zoonoses Data Collection for providing the original datasets necessary to conduct this study. The EU model was developed with funding from contract CT/EFSA/Zoonoses/2010/02 between EFSA and the DTU National Food Institute, in relation to Question n° EFSA-Q-2010-00685.

REFERENCES

1. World Health Organization. The world health report 2007. (http://www.who.int/whr/2007/en/index.html). Geneva, Switzerland: World Health Organization. Acessed 7 July 2011.
2. Scallan E, *et al*. Foodborne illness acquired in the United States— major pathogens. Emerging Infectious Diseases 2011; 17: 7-15.
3. Hendriksen RS, *et al.* Global monitoring of *Salmonella* serovar distribution from the World Health Organization Global Foodborne Infections Network Country Data Bank: results of quality assured laboratories from 2001 to 2007. Foodborne Pathogens and Disease 2011; 8: 887-900.
4. Majowicz SE, *et al.* The global burden of nontyphoidal *Salmonella* gastroenteritis. Clinical Infectious Diseases 2010; 50: 882-889.

5. European Food Safety Authority. The community summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in the European Union in 2008. EFSA Journal 2010; 8: 23-110.

6. Baker MG, *et al*. A recurring salmonellosis epidemic in New Zealand linked to contact with sheep. Epidemiology and Infection 2007; 135: 76-83.

7. O'Reilly CE, *et al*. A waterborne outbreak of gastroenteritis with multiple etiologies among resort island visitors and residents: Ohio, 2004. Clinical Infectious Diseases 2007; 44: 506-512.

8. Pires SM, *et al*. Attributing the human disease burden of foodborne infections to specific sources. Foodborne Pathogens and Disease 2009; 6: 417-424.

9. Havelaar A, *et al*. Towards an integrated approach in supporting microbiological food safety decisions. Zoonosis and Public Health 2007; 54: 103-107.

10. Hald T, *et al*. A Bayesian approach to quantify the contribution of animal-food sources to human salmonellosis. Risk Analysis 2004; 24: 255-269.

11. Pires SM, Hald T. Assessing the differences in public-health impact of *Salmonella* subtypes using a Bayesian microbial subtyping approach for source attribution. Foodborne Pathogens and Diseases 2010; 7: 143-151.

12. Mullner P, *et al*. Source attribution of food-borne zoonoses in New Zealand: a modified Hald model. Risk Analysis 2009; 29: 970-984.

13. Whalström H, *et al.* Source attribution of human *Salmonella* cases in Sweden. Epidemiology and Infection 2011; 139: 1246-1253.

14. Valkenburgh S, *et al*. Zoonoses and zoonotic agents in humans, food, animals and feed in the Netherlands 2003–2006. Bilthoven, the Netherlands: Rijksinstituut voor Volksgezondheid en Milieu, 2007. RIVM rapportnummer: 330152001. ISBN-13, 978-90-6960-184-7.

15. Pires SM, *et al.* Using outbreak data for source attribution of human salmonellosis and campylobacteriosis in Europe. Foodborne Pathogens and Disease 2010; 7: 1351-1361.

16. Statistical Office of the European Union (EUROSTAT). (epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database). Acessed 25 March 2011.

17. European Food Safety Authority. Report on the analysis of the baseline study on the prevalence of *Salmonella* in holdings of laying hen flocks of *Gallus gallus*. EFSA Journal 2007; 97:1-84.

18. European Food Safety Authority. Report of the task force on zoonoses data collection on the analysis of the baseline survey on the prevalence of *Salmonella* in turkey flocks, in the EU, 2006-2007. Part A: *Salmonella* prevalence estimates..The EFSA Journal 2008; 134:1-224.

19. European Food Safety Authority. Report of the task force on zoonoses data collection on the analysis of the baseline survey on the prevalence of *Salmonella* in slaughter pigs in the EU, 2006-2007. Part A: *Salmonella* prevalence estimates. The EFSA Journal 2008; 135:1-111.

20. European Food Safety Authority. Analysis of the baseline survey on the prevalence of *Campylobacter* in broiler batches and of *Campylobacter* and *Salmonella* on broiler carcasses in the EU, 2008 - Part A: *Campylobacter* and *Salmonella* prevalence estimates. EFSA Journal 2010; 8:1503.

21. European Food Safety Authority. The community summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in the European Union in 2006. EFSA Journal 2007; 130:352.

22. European Food Safety Authority. The community summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in the European Union in 2007. EFSA Journal 2009; 223:312.

23. European Food Safety Authority. The community summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in the European Union in 2008. EFSA Journal 2010; 8:1496.

24. European Food Safety Authority. The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2009. EFSA Journal 2011; 9:2090.

25. de Jong B, Ekdahl K. The comparative burden of salmonellosis in the European Union member states, associated and candidate countries. BMC Public Health 2006; 6:4.

26. European Centre for Disease Prevention and Control. 2007. Surveillance of communicable diseases in the European Union: A long-term strategy 2008–2013. (http://ecdc.europa.eu/en/aboutus/Key%20Documents/08-13_KD_Surveillance_of_CD.pdf). Solna, Sweden: European Centre for Disease Prevention and Control, 2007. Acessed 02 November 2012.

27. Havelaar AH, *et al*. Estimating the true incidence of campylobacteriosis and salmonellosis in the European Union, 2009. Epidemiology and Infection 2012; 141: 293-302.

28. Hald T, Pires SM, de Knegt LV. Development of a *Salmonella* source-attribution model for evaluating targets in the turkey meat production. External scientific report submitted to EFSA. Supporting Publications 2012:EN-259.

29. World Health Organization Collaborating Center for Reference and Research on *Salmonella*. Antigenic formulae of the *Salmonella* serovars. 9th Edition. Paris, France: Institut Pasteur, 2007.

30. World Health Organization Global Foodborne Infections Network Country Databank (WHO/GFN CDB). (http://thor.dfvf.dk/gss) Acessed 15 November 2010.

31. European Food Safety Authority. Panel on biological hazards (BIOHAZ); Scientific opinion on monitoring and assessment of the public health risk of "*Salmonella* Typhimurium-like" strains. EFSA Journal 2010; 8:1826.

32. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. Statistical Science 1992; 7: 457-472.

33. Anon. Annual report on zoonoses in Denmark 2009. Copenhagen, Denmark: The Ministry of Food, Agriculture and Fisheries, 2010.

34. Anon. Annual report on zoonoses in Denmark 1998. Copenhagen, Denmark: The Ministry of Food, Agriculture and Fisheries, 1999.

35. Wegener HC *et al*. *Salmonella* control programs in Denmark. Emerging Infectious Diseases 2003; 9:774-780.

Tables

Table 1. Human cases of salmonellosis reported in the modeling dataset before and after adjusting for underreporting (UFs with 95% Credibility Intervals).

Country	Reported	UF (95% CI)[(a)]	Adjusted
AT	8,487	11 (1.6 , 33.6)	93,357 (13,579 , 285,163)
BE	11,066	3.5 (0.3 , 12.5)	38,731 (3,320 , 138,325)
BG	3,899	718.5 (112 , 2141)	2,801,432 (435,518 , 8,345,810)
CY	471	173.2 (26.8 , 523.8)	81,577 (12,623 , 246,710)
CZ	38,842	28.9 (4.3 , 86)	1,122,534 (167,021 , 3,340,412)
DE	127,330	9.8 (1.5 , 29.3)	1,247,834 (190,995 , 3,730,769)
DK	7,497	4.4 (0.7 , 13.1)	32,987 (5,248 , 98,211)
EE	1,341	16.9 (2.4 , 51.8)	22,663 (3,218 , 69,464)
ES	12,033	214.2 (32.7 , 638.9)	2,577,469 (393,479 , 7,687,884)
FI	8,228	0.4 (0 , 1.2)	3,291 (0 , 9,874)
FR	20,319	26.9 (4 , 82)	546,581 (81,276 , 1,666,158)
GR	1,927	1228.5 (189 , 3668)	2,367,320 (363,240 , 7,068,621)
HU	19,091	66.8 (10.2 , 199.1)	1,275,279 (194,728 , 3,801,018)
IE	1,264	5.4 (0 , 27.2)	6,826 (0 , 34,381)
IT	10,205	71.7 (10.7 , 214)	731,699 (109,194 , 2,183,870)
LT	7,643	59.1 (8.7 , 182.1)	451,701 (66,494 , 1,391,790)
LU	479	4.5 (0 , 21.4)	2,156 (0 , 10,251)
LV	2,665	43.3 (6.6 , 134.9)	115,395 (17,589 , 359,509)
MT	371	222.7 (33.7 , 663)	82,622 (12,503 , 245,973)
NL	4,168	26.3 (3.6 , 84.8)	109,618 (15,005 , 353,446)
PL	30,963	114.1 (17.2 , 338.2)	3,532,878 (532,564 , 10,471,687)
PT	1,513	2082.9 (318 , 6267)	3,151,428 (481,588 , 9,481,820)
RO	2,351	349.9 (48 , 1128)	822,615 (112,848 , 2,651,458)
SE	11,265	40.3 (4.9 , 133.2)	453,980 (55,199 , 1,500,498)
SE	3,002	0.5 (0.1 , 1.6)	1,501 (300 , 4,803)
SK	19,399	53.2 (7.6 , 165.4)	1,032,027 (147,432 , 3,208,595)
UK	36,666	7.3 (1.1 , 22.6)	267,662 (40,333 , 828,652)
EU-27	392,485	57.5 (8.8 , 171.4)	22,567,888 (3,453,868 , 67,271,929)

Table 2. Parameters used to estimate the number of sporadic cases of salmonellosis attributable to the animal sources

Notation	Description	Estimation
i (1-22)	*Salmonella* serovar	-
j (1-4)	Food-animal source	
c (1-24)	Country where the human case was reported	
k (1-24)	Country of origin of the food product[(a)]	
o_{ci}	Observed cases caused by serovar i in country c	Data
ob_{ci}	Observed cases caused by serovar i known to be outbreak related in country c. For each outbreak, one case was subtracted so that one outbreak contributed with one sporadic case.	Data
yt_{ci}	Observed cases caused by serovar i in country c that was reported as travel-related	Data
p_{kji}	Prevalence of serovar i in source j in country k	Data
m_{ckj}	Amount of source j available for consumption in country c produced in country k[(a)]	Data
a_{cj}	Source-dependent factor for source j and country c	dunif(0,max a_{cj})
q_i	Subtype-dependent factor for serovar i	dunif(0,max q_i)
uf_c	Underreporting factor for country c	dllnorm(μ,σ)
$spdo_{ci}$	Total number of sporadic cases caused by serovar i in country c	$o_{ci} - yt_{ci} - (ob_{ci} + 1)$

(a)If the food is produced and consumed in the same country, c=k

Table 3. Estimated proportion of human reported cases by food-animal source and the top-5 serovars within each source.

Animal source associated to cases							
Broilers		**Layers**		**Pigs**		**Turkeys**	
Serovar	**%**	**Serovar**	**%**	**Serovar**	**%**	**Serovar**	**%**
Enteritidis	85.0	Enteritidis	95.0	Typhimurium	50.9	Enteritidis	27.9
Infantis	4.5	Typhimurium	1.4	Enteritidis	38.2	Typhimurium	18.6
Typhimurium	2.5	Infantis	1.3	Derby	1.8	Newport	9.2
Virchow	2.9	Virchow	1.0	Infantis	1.1	Saintpaul	4.5
Kentucky	0.6	Kentucky	0.2	Newport	2.3	Hadar	19.0
Others	4.5	Others	1.0	Others	5.7	Others	21.0
Total cases	2,348,384	Total cases	7,899,435	Total cases	5,789,456	Total cases	702,335

Figures

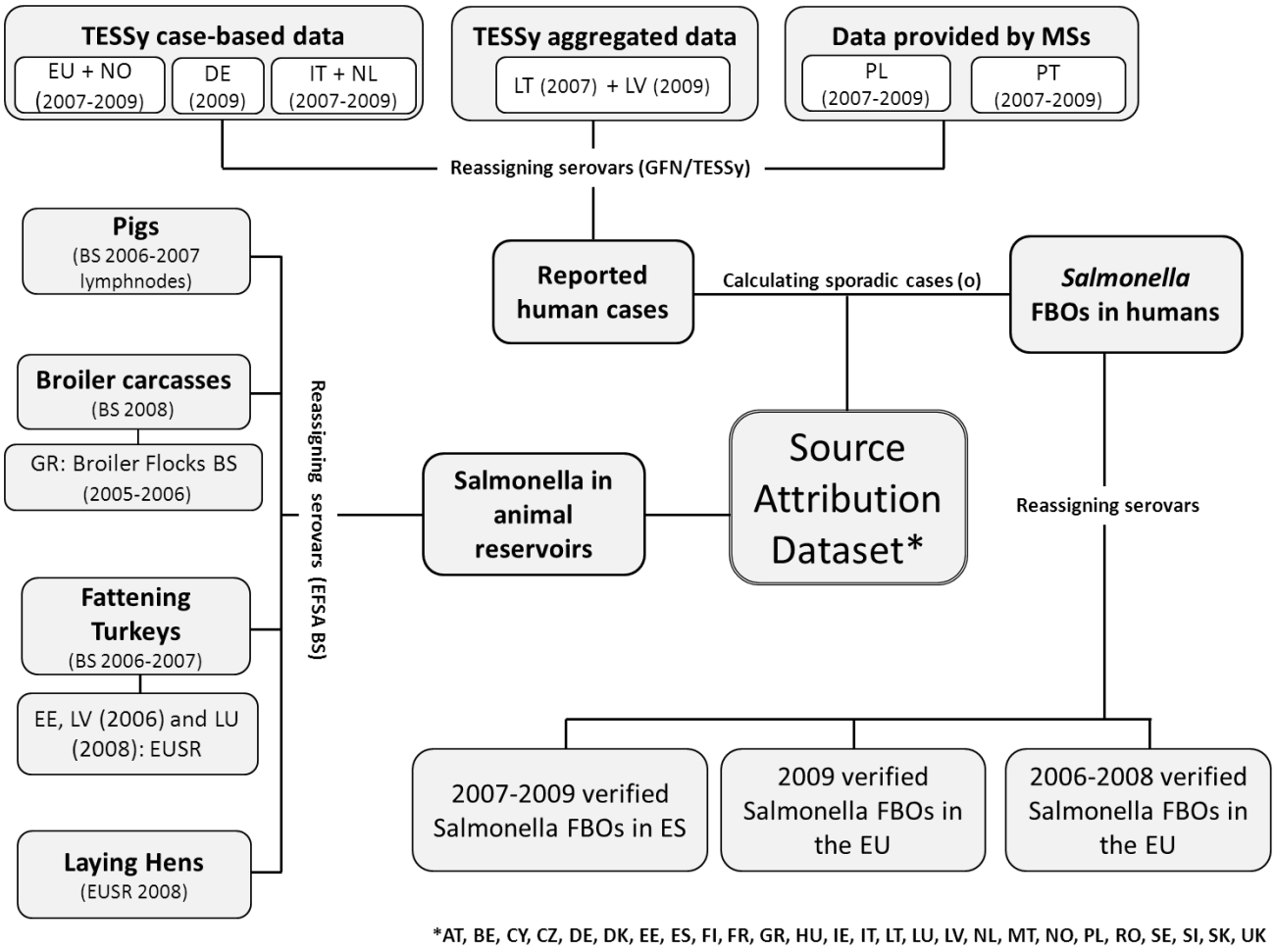


Figure 1. Diagram illustrating the construction of the final dataset for source attribution. For animal reservoirs and outbreaks, each gray block represents a dataset. For reported human cases, white blocks represent primary datasets originally provided to compose the gray blocks.

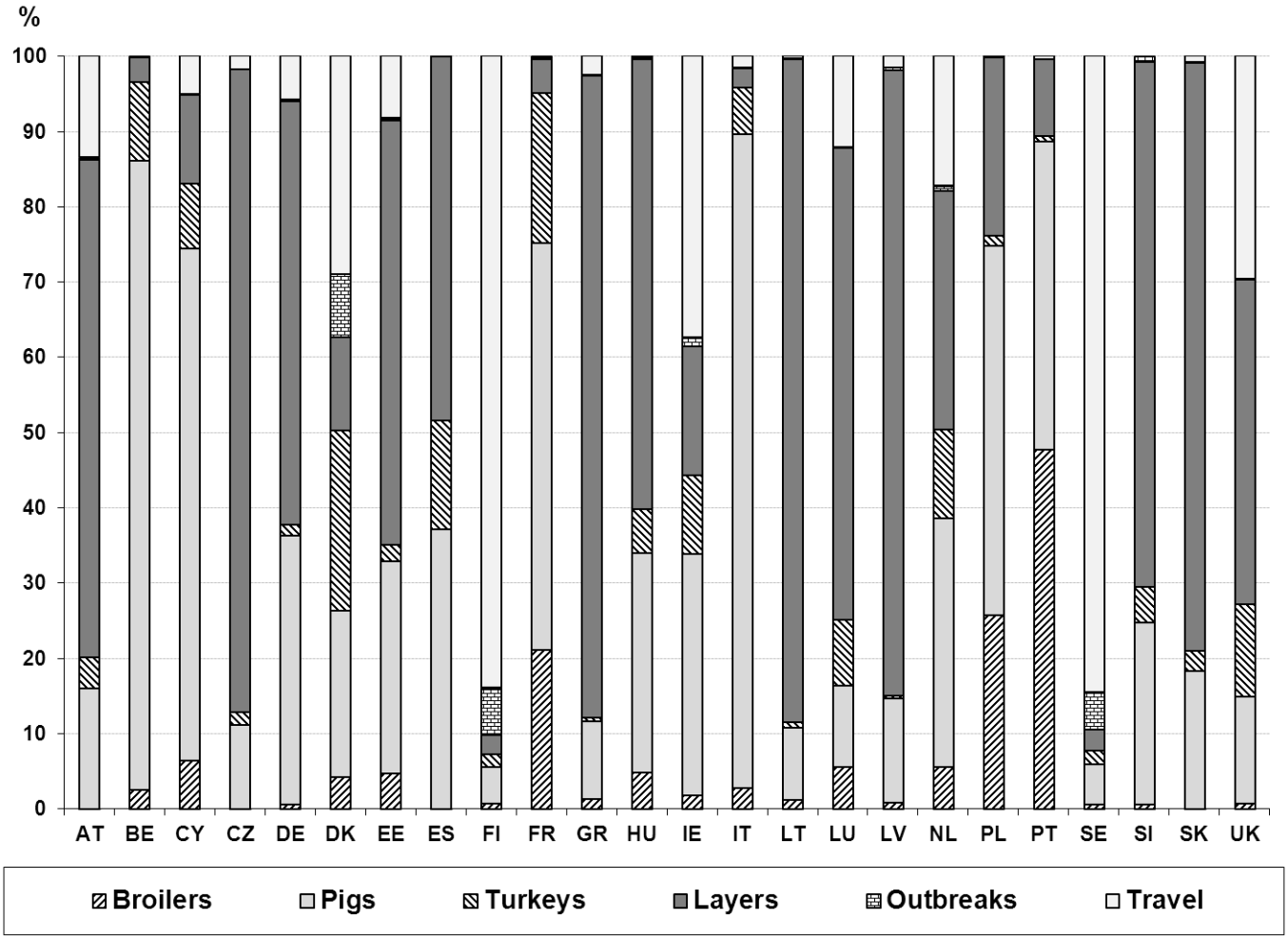


Figure 2. Proportion of *Salmonella* human cases attributed to food animal reservoirs, travel and outbreaks in 24 EU Member States, 2007-2009 (mean %).

Appendix A. Country-specific attribution estimates to food-animal reservoirs, travel, outbreaks and unknown sources.

Source	AT				BE				CY			
	%	mean	95% CI		%	mean	95% CI		%	mean	95% CI	
Broilers	0.1	73	26	277	2.3	935	104	3,672	4.8	4,226	544	15,500
Pigs	14.4	13,130	1,971	45,970	74.2	30,130	3,461	117,300	51.1	44,580	6,639	156,700
Turkeys	3.7	3,417	503	12,090	9.2	3,750	423	14,680	6.4	5,626	618	21,480
Layers	59.8	54,520	8,310	189,500	2.9	1,178	123	4,710	8.9	7,722	976	28,520
Travel	12.2	11,110	1,674	38,690	0.0	0	0	0	3.8	3,334	504	11,650
Unknown	9.4	8,605	1,267	30,210	11.2	4,554	512	17,810	24.9	21,750	3,128	77,430
Outbreak	0.3	272			0.1	52			0.0	0		

Source	CZ				DE				DK			
	%	mean	95% CI		%	mean	95% CI		%	mean	95% CI	
Broilers	0.1	1,308	92	5,201	0.5	6,378	519	24,980	3.5	918	132	3,295
Pigs	10.9	128,900	19,490	446,700	33.1	420,300	63,750	1,462,000	18.0	4,743	854	16,170
Turkeys	1.8	20,710	3,080	72,250	1.3	17,000	2,561	59,330	19.6	5,167	775	18,210
Layers	84.6	997,000	151,300	3,450,000	52.0	660,800	100,100	2,301,000	10.1	2,665	617	8,710
Travel	1.7	20,090	3,047	69,610	5.3	67,860	10,260	236,100	23.7	6,239	946	21,850
Unknown	0.8	9,890	-1,204	41,970	7.6	96,850	14,570	337,000	18.3	4,813	725	16,860
Outbreak	0.0	88			0.2	1,990			6.8	1,786		

Source	EE				ES				FI			
	%	mean	95% CI		%	mean	95% CI		%	mean	95% CI	
Broilers	4.6	923	160	3,356	0.1	3,384	45	17,680	0.7	21	0	96
Pigs	27.5	5,488	818	19,130	33.1	869,600	130,000	3,066,000	4.7	150	22	530
Turkeys	2.1	421	47	1,601	12.9	339,100	50,400	1,196,000	1.6	53	5	203
Layers	55.0	10,980	1,671	37,940	43.1	1,133,000	169,200	4,003,000	2.4	79	10	291
Travel	7.9	1,587	244	5,460	0.0	0	0	0	80.1	2,571	387	8,939
Unknown	2.6	516	-601	2,764	10.7	281,100	41,470	993,700	4.6	148	21	530
Outbreak	0.3	63			0.0	469			5.9	189		

Source	FR				GR				HU			
	%	mean	95% CI		%	mean	95% CI		%	mean	95% CI	
Broilers	13.4	66,000	10,120	230,000	1.2	28,530	384	148,100	4.5	52,570	7,904	182,900
Pigs	34.3	168,900	25,950	586,700	9.5	227,200	33,520	801,600	26.7	313,300	47,160	1,090,000
Turkeys	12.6	62,180	9,363	217,400	0.4	9,061	468	40,570	5.4	63,760	9,558	222,200
Layers	2.9	14,150	2,864	47,600	78.3	1,872,000	279,200	6,552,000	54.9	643,600	96,960	2,231,000
Travel	0.0	0	0	0	2.3	55,820	8,336	195,400	0.2	1,975	298	6,840
Unknown	36.5	179,800	27,140	627,000	8.3	197,700	25,090	721,300	8.1	94,870	14,110	331,500
Outbreak	0.2	966			0.0	0			0.2	1,815		

Source	IE				IT				LT			
	%	mean	95% CI		%	mean	95% CI		%	mean	95% CI	
Broilers	1.5	100	6	486	2.3	17,680	2,639	62,030	1.2	5,244	631	19,630
Pigs	27.2	1,810	113	8,616	73.2	560,700	85,200	1,949,000	9.5	42,750	6,428	148,700
Turkeys	8.8	589	35	2,810	5.3	40,410	6,028	141,700	0.7	3,318	369	12,600
Layers	14.6	971	64	4,594	2.2	16,520	2,309	59,450	86.9	390,000	59,010	1,353,000
Travel	31.7	2,110	133	10,020	1.3	9,908	1,505	34,480	0.3	1,294	196	4,488
Unknown	15.3	1,018	62	4,864	15.8	120,800	18,280	421,300	1.2	5,596	-4,449	27,910
Outbreak	0.9	63			0.0	0			0.1	335		

Source	LU				LV				NL			
	%	mean	95% CI		%	mean	95% CI		%	mean	95% CI	
Broilers	4.4	96	6	449	0.9	873	92	4,135	4.6	4,455	653	15,810
Pigs	8.5	184	13	833	13.7	13,590	2,052	47,280	27.3	26,330	3,978	91,590
Turkeys	6.9	149	11	670	0.3	291	6	1,368	9.7	9,404	1,393	33,050
Layers	49.8	1,073	89	4,662	82.5	81,600	12,450	282,200	26.2	25,270	4,015	87,770
Travel	9.6	207	17	896	1.5	1,459	222	5,046	14.2	13,730	2,079	47,900
Unknown	20.7	446	35	1,961	0.7	714	-4,236	7,337	17.5	16,920	2,521	59,240
Outbreak	0.0	0			0.4	351			0.5	470		

Source	PL				PT				SE			
	%	mean	95% CI		%	mean	95% CI		%	mean	95% CI	
Broilers	25.1	796,600	120,900	2,772,000	42.3	1,357,000	202,700	4,727,000	0.5	28	2	117
Pigs	47.8	1,520,000	229,700	5,269,000	36.3	1,164,000	175,500	4,052,000	4.8	282	42	991
Turkeys	1.2	39,640	5,790	139,900	0.6	18,580	546	83,890	1.7	99	13	361
Layers	23.0	731,300	111,500	2,550,000	9.1	290,400	29,270	1,138,000	2.5	145	29	506
Travel	0.1	1,978	300	6,882	0.4	11,250	1,704	39,030	75.9	4,441	666	15,530
Unknown	2.7	84,840	11,520	305,300	11.4	364,300	49,970	1,310,000	10.2	596	89	2,089
Outbreak	0.1	3,484			0.0	90			4.4	260		

Source	SI				SK				UK			
	%	mean	95% CI		%	mean	95% CI		%	mean	95% CI	
Broilers	0.5	564	7	3,037	0.0	363	21	1,779	0.6	1,590	236	5,565
Pigs	20.6	21,600	2,464	84,410	18.0	189,300	28,490	664,900	11.7	32,370	4,886	112,600
Turkeys	4.0	4,197	452	16,740	2.6	27,580	4,066	97,930	10.1	27,930	4,208	97,290
Layers	59.5	62,240	7,195	242,500	76.8	807,500	121,800	2,826,000	35.5	97,990	14,800	340,900
Travel	0.0	0	0	0	0.8	8,152	1,228	28,540	24.3	67,250	10,170	234,100
Unknown	14.7	15,370	1,716	60,450	1.7	17,940	1,124	70,500	17.8	49,270	7,449	171,200
Outbreak	0.6	656			0.0	449			0.0	0		

Appendix B. Attribution estimates to food-animal reservoirs in their country of origin. The percentage column refers to percentage of EU cases "originated" by that country.

	AT				BE				CY		
Source	**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**	
Broilers	0.3	191	67	466	0.8	16,540	5,603	44,550	0.3	4,155	552	15,190
Pigs		23,560	10,910	47,650		109,000	53,050	220,300		40,090	10,190	122,900
Turkeys		2,437	947	5,810		851	258	2,516		456	50	1,741
Layers		31,970	6,051	107,300		14,340	3,923	43,140		3,045	386	11,220

	CZ				DE				DK		
Source	**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**	
Broilers	6.0	956	247	3,045	6.7	7,650	2,587	20,900	0.5	0	0	0
Pigs		114,700	29,510	323,100		645,100	265,700	1,551,000		85,460	37,260	189,400
Turkeys		15,020	4,451	40,000		22,310	10,170	50,260		0	0	0
Layers		874,200	142,000	2,999,000		440,100	124,900	1,258,000		584	226	1,390

	EE				ES				FI		
Source	**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**	
Broilers	0.1	0	0	0	18.4	58,490	13,020	185,800	0.0	0	0	0
Pigs		9,121	3,271	21,790		1,306,000	423,700	3,556,000		0	0	0
Turkeys		0	0	0		302,600	55,350	1,029,000		0	0	0
Layers		5,419	1,339	16,020		1,414,000	406,100	4,286,000		10	4	22

	FR				GR				HU			
Source	**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**	
Broilers	2.5	35,210	6,162	120,300	10.9	27,380	499	141,600	6.0	53,660	9,881	180,800
Pigs		238,400	91,980	576,800		90,550	13,560	319,100		286,600	59,800	943,700
Turkeys		116,700	43,460	287,300		445	54	1,754		84,060	27,580	230,500
Layers		20,610	7,262	52,790		1,701,000	256,400	5,944,000		587,900	93,970	2,023,000

	IE				IT				LT			
Source	**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**	
Broilers	0.1	3,927	1,474	9,996	2.4	17,440	3,194	59,030	0.0	1,192	216	4,185
Pigs		8,004	4,158	15,200		299,900	51,940	1,024,000		4,684	791	16,020
Turkeys		638	185	1,809		56,860	19,810	153,800		399	108	1,207
Layers		7	1	21		32,510	10,850	82,980		0	0	0

	LU				LV				NL			
Source	**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**	
Broilers	0.0	0	0	0	1.2	876	157	3,126	1.8	1,944	890	4,067
Pigs		340	146	785		3,016	544	10,190		121,000	56,900	251,200
Turkeys		0	0	0		0	0	0		5,088	2,711	9,397
Layers		414	39	1,776		196,500	69,420	486,500		165,200	39,940	512,100

	PL				PT				SE			
Source	**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**	
Broilers	21.3	803,600	131,400	2,768,000	14.5	1,305,000	198,500	4,535,000	0.0	7	2	20
Pigs		1,402,000	257,400	4,721,000		876,000	134,800	3,040,000		364	189	695
Turkeys		71,110	30,950	167,900		1,342	198	5,397		0	0	0
Layers		1,287,000	492,000	3,162,000		239,800	27,870	928,000		64	13	215

	SI				SK				UK			
Source	**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**	
Broilers	0.4	412	33	2,060	5.0	1,711	628	3,999	0.8	8,632	3,920	18,300
Pigs		11,440	1,577	43,820		72,300	12,190	249,600		50,810	20,800	117,600
Turkeys		2,864	381	11,170		71	18	220		19,080	6,737	52,040
Layers		57,020	6,929	221,100		768,300	192,800	2,339,000		60,270	10,210	206,300

Appendix D. Estimated values for a_{cj}, source-dependent factor (mean and 95% Credibility Interval).

Country	Broilers	95% CI			Pigs	95% CI			Turkeys	95% CI			Layers	95% CI		
AT	5.38E-07	[2.00E-08	,	2.85E-06]	4.93E-05	[4.30E-05	,	5.58E-05]	1.39E-04	[1.07E-04	,	1.76E-04]	7.42E-04	[7.14E-04	,	7.70E-04]
BE	2.69E-05	[2.10E-05	,	3.36E-05]	2.00E-04	[1.89E-04	,	2.11E-04]	5.27E-04	[4.54E-04	,	6.06E-04]	1.73E-05	[1.02E-05	,	2.48E-05]
CY	1.66E-06	[8.27E-07	,	2.74E-06]	3.51E-05	[2.93E-05	,	4.12E-05]	1.39E-03	[4.90E-04	,	2.59E-03]	1.03E-04	[4.97E-05	,	1.73E-04]
CZ	1.81E-06	[3.07E-07	,	3.86E-06]	1.41E-04	[1.30E-04	,	1.52E-04]	6.25E-04	[4.98E-04	,	7.67E-04]	8.63E-04	[8.52E-04	,	8.74E-04]
DE	1.11E-05	[2.36E-06	,	2.13E-05]	1.13E-04	[1.07E-04	,	1.19E-04]	1.05E-04	[8.92E-05	,	1.22E-04]	6.08E-04	[6.02E-04	,	6.15E-04]
DK	9.31E-05	[5.06E-05	,	1.42E-04]	2.19E-05	[1.91E-05	,	2.48E-05]	2.39E-03	[2.01E-03	,	2.79E-03]	1.08E-04	[8.54E-05	,	1.31E-04]
EE	7.53E-05	[1.54E-05	,	1.55E-04]	9.84E-05	[7.51E-05	,	1.24E-04]	3.43E-04	[1.27E-04	,	6.85E-04]	4.48E-04	[3.92E-04	,	5.04E-04]
ES	4.12E-08	[1.54E-09	,	2.18E-07]	1.07E-05	[9.91E-06	,	1.15E-05]	1.54E-04	[1.33E-04	,	1.76E-04]	1.01E-05	[9.78E-06	,	1.04E-05]
FI	6.27E-04	[3.15E-05	,	1.98E-03]	4.70E-04	[3.35E-04	,	5.98E-04]	2.71E-03	[7.36E-04	,	4.99E-03]	8.39E-06	[3.53E-06	,	1.34E-05]
FR	5.10E-05	[4.73E-05	,	5.50E-05]	3.86E-05	[3.61E-05	,	4.10E-05]	1.47E-04	[1.33E-04	,	1.62E-04]	5.99E-06	[4.65E-06	,	7.50E-06]
GR	6.83E-07	[2.51E-08	,	3.53E-06]	8.43E-06	[6.90E-06	,	1.01E-05]	1.07E-05	[1.51E-06	,	3.47E-05]	2.61E-05	[2.45E-05	,	2.75E-05]
HU	4.93E-06	[4.23E-06	,	5.67E-06]	1.06E-04	[9.72E-05	,	1.16E-04]	2.00E-04	[1.75E-04	,	2.28E-04]	2.39E-04	[2.30E-04	,	2.47E-04]
IE	3.39E-07	[1.83E-07	,	5.52E-07]	2.24E-05	[1.88E-05	,	2.61E-05]	2.34E-04	[1.56E-04	,	3.26E-04]	3.95E-05	[3.04E-05	,	4.91E-05]
IT	6.56E-06	[5.22E-06	,	8.06E-06]	5.81E-05	[5.51E-05	,	6.11E-05]	4.52E-05	[3.52E-05	,	5.72E-05]	1.34E-06	[8.14E-07	,	2.00E-06]
LT	2.08E-05	[7.6E-06	,	3.5E-05]	1.2E-04	[1.0E-04	,	1.4E-04]	1.3E-04	[4.6E-05	,	2.5E-04]	3.8E-02	[3.7E-02	,	3.9E-02]
LU	3.11E-05	[5.6E-06	,	5.2E-05]	2.8E-05	[1.2E-05	,	5.0E-05]	4.0E-04	[1.9E-04	,	6.9E-04]	6.8E-04	[5.5E-04	,	8.1E-04]
LV	2.85E-06	[9.1E-08	,	1.2E-05]	7.5E-05	[6.0E-05	,	9.2E-05]	7.5E-05	[3.2E-06	,	3.0E-04]	9.4E-05	[8.8E-05	,	1.0E-04]
NL	5.37E-06	[3.7E-06	,	6.7E-06]	2.0E-05	[1.8E-05	,	2.2E-05]	1.5E-04	[1.1E-04	,	1.8E-04]	2.9E-05	[2.6E-05	,	3.1E-05]
PL	2.01E-05	[1.7E-05	,	2.0E-05]	5.8E-05	[5.4E-05	,	6.3E-05]	2.1E-05	[1.5E-05	,	2.9E-05]	3.9E-05	[3.4E-05	,	4.5E-05]
PT	8.37E-06	[6.8E+03	,	9.8E-06]	8.2E-06	[7.3E-06	,	9.1E-06]	5.4E-06	[2.9E-07	,	1.8E-05]	2.3E-06	[7.4E-07	,	4.9E-06]
SE	1.46E-04	[2.4E-05	,	2.4E-04]	7.7E-05	[6.3E-05	,	9.1E-05]	5.0E-03	[3.1E-03	,	7.1E-03]	3.8E-04	[2.3E-04	,	5.4E-04]
SI	6.52E-06	[2.2E-07	,	2.9E-05]	1.3E-04	[1.1E-04	,	1.5E-04]	8.8E-05	[5.5E-05	,	1.3E-04]	2.3E-04	[2.1E-04	,	2.4E-04]
SK	1.50E-07	[5.5E-09	,	6.6E-07]	3.8E-04	[3.5E-04	,	4.1E-04]	6.4E-04	[5.0E-04	,	8.0E-04]	1.1E-03	[1.0E-03	,	1.1E-03]
UK	1.71E-06	[1.1E-06	,	2.0E-06]	4.4E-05	[4.0E-05	,	4.8E-05]	1.2E-03	[1.1E-03	,	1.4E-03]	4.8E-04	[4.7E-04	,	4.9E-04]

Appendix E. Estimated values for *qi*, *Salmonella* subtype-dependent factor (mean and 95% Credibility Interval).

Serovar	***qi***	**95% CI**
***S*. Enteritidis**	1[(a)]	
***S*. Agona**	0.0527	[0.0488 , 0.0569]
***S*. Anatum**	0.0252	[0.0223 , 0.0283]
***S*. Bovismorbificans**	0.1854	[0.1690 , 0.2034]
***S*. Brænderup**	0.1386	[0.1223 , 0.1567]
***S*. Brandenburg**	0.1096	[0.1009 , 0.1190]
***S*. Bredeney**	0.0170	[0.0151 , 0.0191]
***S*. Derby**	0.0197	[0.0186 , 0.0201]
***S*. Hadar**	0.0734	[0.0670 , 0.0806]
***S*. Heidelberg**	0.1163	[0.0960 , 0.1401]
***S*. Infantis**	0.1223	[0.1167 , 0.1281]
***S*. Kentucky**	1.9980	[1.7970 , 2.2130]
***S*. Kottbus**	0.0143	[0.0124 , 0.0164]
***S*. Livingstone**	0.0595	[0.0540 , 0.0653]
***S*. London**	0.0826	[0.0751 , 0.0908]
***S*. Mbandaka**	0.0473	[0.0425 , 0.0523]
***S*. Montevideo**	0.1124	[0.1044 , 0.1210]
***S*. Newport**	0.2476	[0.2320 , 0.2645]
***S*. Rissen**	0.0302	[0.0268 , 0.0340]
***S*. Saintpaul**	0.0600	[0.0538 , 0.0671]
***S*. Typhimurium**	0.2153	[0.2054 , 0.2264]
***S*. Virchow**	0.2469	[0.2320 , 0.2625]

(a) The q value for *S*. Enteritidis is fixed to 1, and the other serovars are calculated relatively to it.

Manuscript III

Sources of human salmonellosis in Denmark: comparing the results of the Danish *Salmonella* source account model with a source attribution model developed at EU level

L.V. DE KNEGT[(1)], S.M. PIRES[(1)], T. HALD[(1)]

[(1)],National Food Institute, Division of Epidemiology and Microbial Genomics

Technical University of Denmark

Author responsible for manuscript and reprint correspondence: L.V. DE KNEGT

Mailing address:

Technical University of Denmark, National Food Institute

Division of Epidemiology and Microbial Genomics

Mørkhøj Bygade 19, Building H, 2860 Søborg, Denmark

Email: ledkn@food.dtu.dk

Running head: Comparing source attribution in DK and the EU

Submitted to Foodborne Pathogens and Disease

Abstract

Danish risk management strategies for *Salmonella* control in the food chain rely on the routine application of a source attribution model to estimate the contribution of the major animal-food sources to human infections. This model was the basis for the development of a European Union model, which needed to be validated. As part of this process, results obtained for Denmark by the EU model were compared with the ones obtained using the Danish model in the same period. The latter estimated pigs as the main animal source of human salmonellosis in the period between 2007 and 2009, followed by table eggs (layers) and broilers, while in the EU model the estimated order of importance in the same period was turkeys, pigs, layers and broilers. Differences in travel-related cases and cases that could not be attributed to any source were also observed. Discrepancies in numbers are mostly explained by differences in the management of missing data, the level of subtyping available in the two datasets, the number of sources included in each model, the inclusion of multiple countries, and the use of trade data to estimate cases caused by imported food. Still, the two models ranked three out of the four sources in a similar order. We concluded that the EU model is useful for countries which cannot readily attain the level of data detailing achieved by Danish monitoring and surveillance, and that the Danish model would benefit more from adapting the EU model's approach to use food trade data.

Key words: source attribution; source account; *Salmonella*

INTRODUCTION

During the last decades, *Salmonella* has been one of the major causes of foodborne infections in Denmark. Different food animals have been identified as the main reservoir in different moments, such as broilers in the late eighties, pigs in the early nineties and laying hens in the late nineties and 2000's (Wegener 2010). For that reason, Denmark has developed a series of initiatives to control and reduce the spread of *Salmonella* in the food chain, which rely on a well-established system for *Salmonella* surveillance in humans, foods and animals.

As part of the risk-management activities, the National Food Institute routinely applies a source attribution model to estimate the contribution of the major animal-food sources to human infections of *Salmonella*. The model was first implemented in 1995 and has been evolving ever since, becoming a stochastic model in 2004 (Hald et al., 2004), and developing the possibility of using data from multiple years in 2008, (Pires and Hald, 2010), thereby obtaining more robust and accurate results..

The principle of source attribution by microbial subtyping is to compare the occurrence of subtypes in animals or food sources with the same subtypes in humans. The approach is enabled by the identification of strong associations between some of the dominant subtypes and a specific reservoir or source, as long as a heterogeneous distribution of subtypes among the sources exists. Infections by subtypes exclusively or almost exclusively isolated from one source are regarded as originating from that source. Human infections caused by subtypes found in several reservoirs are then distributed relatively to the prevalence of the indicator types. This approach utilizes a collection of temporally- and spatially-related isolates from various sources, and thus is facilitated by an integrated foodborne disease surveillance programme that is focused on the collection of isolates from the major food animal reservoirs of foodborne diseases (Hald et al. 2004).

A EU model was created in 2010 as an answer to EFSA's need of an overview of the main *Salmonella* animal sources in the European Union (Pires et al. 2011). It is based on the Danish model, but instead of using data from multiple years from one single country, it uses data from several countries in a cross-sectional timeframe (de Knegt, 2013).

The objective of this study was to validate the utility of the newly developed EU model, and conclude on advantages and limitations of both models. This was done by comparing the results of the Danish *Salmonella* source account as published in the Annual Report on Zoonoses in Denmark each year with the source attribution results obtained for Denmark in the EU source attribution study conducted in 2011.

MATERIALS AND METHODS

Both the current Danish model and the EU model are based on the original Hald model (Hald et al. 2004), which includes two dimensions: the *Salmonella* subtype and the food/animal source. It attributes *Salmonella* sporadic cases to the sources, to international travel and to an unknown source each year. A sporadic case is defined as a subject that could not be associated with a recognized foodborne disease outbreak. Outbreak-related cases except one per subtype are subtracted from the total number of observed cases. The remaining outbreak-cases are added to the final results of the model and attributed either to the source implicated in the outbreak or to outbreaks with unknown source. A domestic case is defined as a subject who had not been traveling before the disease onset. It is assumed that all cases that had been travelling abroad one week prior to onset of symptoms were travel-related. Not all cases have travel information, and human cases with unknown travel information are attributed to travel based the proportion of on the distribution of travelers and non-travelers for each subtype. Cases that are attributed to an unknown source may include cases caused by sources not included in the model or cases caused by isolates that were not subtyped.

The Hald model takes into account the number of cases caused by a serovar, the prevalence of each serovar in each food source and the relative impact of a set of unknown factors. Those factors were included as multi-parameter priors, and account for the differences in the ability of different subtypes to cause disease and of different sources to act as vehicles for infection. An overview of the model parameterization can be drawn as:

$a_j \sim$ Uniform (0,10)

$q_i \sim$ Uniform (0,0.01)

$\lambda_i \sim$ Poisson (o_i),

$\lambda_{ji} = p_{ij} * m_j * a_j * q_i$

where: 1) λ_{ji} is the expected number of cases per serovar *i* in source *j*; 2) p_{ij} is the prevalence of serovar *i* in source *j*; 3) m_j is the amount of source *j* available for consumption in the country; 4) a_j is the source-dependent factor for source *j*; 5) q_i is the subtype-dependent factor for serovar *i;* The source-dependent factor a_j accounts for variability in surveillance of different food sources, also including general variations between sources, e.g. bacterial load/concentration in the food and processing, handling or preparation practices. The subtype-dependent factor q_i is a one-dimensional parameter, meaning that it is a property of the *Salmonella* serovar and assumed independent of the source where it is found. The q_i prior for *S.* Enteritidis is defined as 1, and all other q_i values are estimated relatively to this one. The number of human sporadic and domestic cases attributed to each subtype (λ_i) is estimated assuming a Poisson distribution of the observed number of sporadic cases per subtype (o_i) (Hald et al., 2004).

The current Danish model has three dimensions: *Salmonella* subtype, food/animal source and year (Pires and Hald, 2010).The EU model, albeit also three-dimensional, uses the *Salmonella* subtype, the food/animal source and the country of reporting (de Knegt 2013). Although both models follow the same mathematical approach used in the Hald model, the difference in dimensions require the use of different types of data, mostly because data availability varies between countries. As an example, the EU model includes data from 24 countries, and available data had only serotyping information and varied in representativeness and quality. On the contrary, the Danish model makes use of data with higher discriminatory power, i.e. with a better resolution of subtyping. Additionally, considering national differences in surveillance systems and the levels of underreporting which are expected to happen in different countries (de Jong and Ekdahl 2006; Havelaar et al. 2012), the EU model makes use of underreporting factors, which are multiplied with the

estimated cases after attribution. For Denmark, this means that the "true" number of cases of salmonellosis in the country is expected to be 4.4 times the total officially reported (Havelaar et al. 2012).

Regarding the prevalence (p), the EU model uses data from the EFSA Baseline Studies (BS) on the prevalence of *Salmonella* in broiler carcasses (EFSA 2010a), slaughter pigs (EFSA 2008a) and turkeys (EFSA 2008b), as well as EU-harmonized surveillance data for laying hens (EFSA 2010b), whereas the Danish model uses data from national monitoring programs and from the Case-by-case Risk Assessment Program (CBC). The national *Salmonella* surveillance programmes collect faecal samples at farm level from layers and broiler flocks, and samples from pigs and cattle carcass swabs at slaughter; the latter are assumed to represent the reservoir (farm) level (Anon., 2011). The CBC started in 2007 and collects samples from batches of Danish and imported pork, beef, chicken, turkey and duck, which are then tested for *Salmonella*. Apart from the CBC, individual retail samples are also collected from domestic and imported ducks and turkeys. All isolates collected as part of the surveillance programmes are submitted to the National Food Institute (Food-DTU) for serotyping, and all *S.* Typhimurium and *S.* Enteritidis are phage typed. Isolates of animal and food origin are tested for antimicrobial susceptibility (Agersø et al. 2011). Results of the CBC *Salmonella* testing are recorded by country of origin, but the prevalence in imported sources used in the model is the overall percentage of positive samples by type of imported meat, regardless of the origin.

In the EU model, the amount of each food source available for consumption (m) was estimated as the amount produced in the country minus what is exported, plus what was imported from each other country (data from the European Statistical Office - EUROSTAT). For the same variable, the Danish model uses the total produced minus exported for domestic sources, and the total imported amount for imported foods, without specifying the origin. This means that the EU model ultimately works in four dimensions, since the country of origin of the food and to which the *Salmonella* prevalences apply can differ from the country where the human cases were reported. Cases are, consequently, attributed to the countries from where the food originated. This combination of data types used for *p* and *m* also means that, while the Danish model uses *m* only for weighting the relative importance of sources, in the EU model the prevalence in a country exporting

a large amount of a food source to Denmark will have a higher impact on the results, when compared to countries exporting smaller amounts. Table 1 presents a summary comparison of the two models.

RESULTS

A total of 7,433 human cases of *Salmonella* were reported in Denmark in the period from 2007 to 2009. Table 2 shows the number of reported cases attributed to animal reservoirs, international travel and outbreaks in each year in the Danish model, as well as the sum of the three years. The most important sources of salmonellosis in this period were pork (7.9% domestic and 1.4% imported), table eggs (7.5%) and broilers (4.7%), of which imported chicken (3.8%) is the largest part. Around 30% of cases were estimated to have been acquired abroad, and 16.7% could not be attributed to any source.

In the EU model, 7,461 cases were reported in Denmark in the same period[1]. After adjusting for underreporting, this resulted in 26,331 cases (Table 3), with turkeys as the most important food source (19.6%), followed by pigs (18.0%), layers (10.1%) and broilers (3.5%). The largest proportion of cases was attributed to international travel (23.7%). Cases that could not be attributed to any source corresponded to 18.3%, and outbreaks with unknown source had 6.8% of cases.

Figure 1 shows the proportion of cases attributable to domestic and imported sources in the two models. The category "others" includes sources present in the Danish model but not in the EU model (e.g. beef and ducks).

DISCUSSION

The comparison between Danish source attribution estimates obtained by the EU *Salmonella* source attribution model and by the Danish single-country model was performed to assess the impact of differences between the two models and conclude on advantages and limitations of each. Differences derive mostly from the type of data used, and reflect the different levels of subtyping, as well as the inclusion of different sources and multiple countries data. The largest discrepancy was observed in the results for turkeys. For

[1] Data sets extracted at different times result in differences in the number of reported cases.

other sources, although individual attributed fractions are different, the order of priority is similar, having pigs as the most important, layers and then broilers (Table 3).

Differences in the data used in the two models explain the discrepancy in the proportion of cases attributed to the turkey reservoir. Both models agreed that all cases attributed to turkeys were related to imported turkey, but the proportion of cases attributed in the EU model was over 10 times the proportion in the Danish model. Table 4 shows that the total amount of turkey meat imported by Denmark as considered in the Danish model was smaller than in the EU model, resulting in a smaller parcel of cases weighted to this reservoir. The data used in the Danish model also shows that the CBC tested samples from the four countries responsible for 88% of cases in the EU model (Germany, France, Italy and Poland), which were also the main exporting countries in the period, according to EUROSTAT. Table 5 shows the prevalence of *S.* Saintpaul, one of the most important turkey serovars, and of *S.* Enteritidis and *S.* Typhimurium, the two most important overall, in Poland, Germany and France (80% of cases). The adjusted number of human cases of each of those serovars in Denmark was 352, 7,044 and 6,310. The prevalence of *S.* Saintpaul in Polish turkeys is 4.3 times the observed in Germany, and over 10 times the French prevalence, while it is almost absent in the other three sources in the three countries. More importantly, Poland is the only country in which the prevalence of *S.* Typhimurium is higher in turkeys than other sources, suggesting that, if more detailed subtyping methods were available, part of the cases attributed to turkeys might have been attributed to other sources (for instance, pigs). Given the large number of *S.*Typhimurium cases (6,310), this results in a large difference, suggesting that the EU model could have overestimated the importance of this source based on Poland, pointed as the main contributor of turkey-originated cases (31%).

Concerning broilers, the absence of cases attributed to domestic broilers in the EU model while the Danish model attributed 3.8% of cases to this source is readily explained by the different data used; in the BS, the prevalence of *Salmonella* in broiler carcasses in Denmark was zero, while the surveillance and monitoring data used in the Danish model had 2.1% positive samples. The parcel attributed to imported broilers in the Danish model is also larger than in the EU model, which is likely explained by the lower level of subtyping detail in the EU model. *S.* Enteritidis is very frequently observed in both broilers and layers, and without

better discriminatory features a parcel of *S.* Enteritidis cases can be misattributed between them. As 27% of sporadic human cases in Denmark (7,044 out of 26,330) were caused by *S.* Enteritidis, this parcel corresponds to a large proportion of total cases. In addition, the Danish model includes data from meat imported from non-EU countries, such as Brazil, Chile and Argentina. The EU model does not take those countries into consideration, which could result in non-EU broiler cases being "forced" into the available countries in this model (Table 6).

The two models tend to agree on the importance pigs for salmonellosis, as 81.5% of cases attributed to this reservoir in the EU model (Table 7) and 84.9% in the Danish model were estimated to be domestic (Table 3). As the overall prevalences in the two largest foreign contributors (Germany and Spain) are similar in the two models (Table 7), this suggests that the main difference in attributable fractions is again likely due to the difference in discriminatory power. *S.* Typhimurium, the most important pig serovar, is one of the main serovars in all other sources, and also the serovar with the second largest amount of human cases (de Knegt 2013). Without better discriminatory power, a large parcel of *S*. Typhimurium cases is attributed to pigs in the EU model, corresponding to a large parcel of total cases. In the Danish model, phage typing data allows better differentiation among sources, resulting in less "cross-attribution".

The different approaches to estimate the contribution of imported foods had an important impact in the results. The EU model used a complex matrix of trade information; as a consequence, when compared to the EU model, the data in the Danish model lead to an underestimation of the contribution of imported meats to human salmonellosis in Denmark. This happens because the CBC does not sample foods from all exporting countries. However, data from the main contributing countries in all categories are available, as well as data from products imported from non-EU countries , suggesting that the CBC data has a good level of sensitivity and representativeness for the purposes of the DK model.

The Danish model estimates the proportion of cases with no travel information that are travel-related, assuming that these should follow the same proportions as the ones for which that information is available; as a consequence, the total cases attributed to travel includes reported and estimated cases. Because no travel

information was available for some countries, cases with no travel information were considered domestic in the EU model, whereas a part of these were attributed to travel in the Danish model. In the EU model dataset, *S*. Enteritidis corresponds to 46% of travel cases in Denmark, where specific phage types of this serovar, like PT 21, PT 4 and PT 6, are most frequently related to travel (Anonymous, 2011). This information cannot be taken into consideration in the EU model, as phage type data were not available. Due to control activities conducted in the country during the last decade, the prevalence of *S*. Enteritidis in food-animals is in general low, particularly when compared to other MS. The EU model, therefore, tends to allocate those cases to countries from which Denmark imports eggs with high *S*. Enteritidis prevalence (since in most MSs eggs are the most common source of *S*. Enteritidis), resulting in another discrepancy: in the EU model, 88% of cases attributed to layers come from imported eggs (Table 8), while the Danish model does not consider this as a valid source. This happens because the *Salmonella* contribution from imported eggs is not considered important in Denmark, as it is believed that most of the imported eggs are not sold as shell eggs, but instead used in heat-treated products. Whether this assumption holds is not known, but the EU model could be improved if this type of country-specific information was used to adjust the results accordingly; in the above example, this would imply in disregarding the cases attributed to imported eggs.

The amount of cases attributed to the "unknown" category is directly affected by the different number of sources in the two models. The attribution of human cases to a limited number of food-animal sources may result in the misplacing of some cases if their "true" source is not included in the model. As an example, it is likely that some beef-related cases in the EU model were "wrongly" attributed to pigs, as *S*. Typhimurium is one of the main serovars in both sources (de Knegt, 2003). The higher level of subtyping detail in the DK model (phage types, AMR profile) also affects this category, since a more discriminatory level of subtyping increases the probability of the model only attributing cases if the right source is included in the model,s, resulting in a larger number of cases being attributed to "unknown". This is also one of the reasons for the animal sources in the EU model receiving in general a larger proportion of cases when compared to the Danish model.

The proportion of cases attributed to outbreaks differed substantially in the two models because the attribution estimates for all sources except outbreaks in the EU model were adjusted for underreporting. It is a model assumption that all outbreak-related cases were properly reported, and so these were not multiplied by the UF. This changed the balance between the proportion of cases attributed to outbreaks and to the other categories, when comparing the two models. If the results of the Danish model are multiplied by 4.4 and the proportions are re-calculated, cases belonging to outbreaks with an unknown source change from 26.4% to 7.6%, which is more similar to the 6.8% estimated by the EU model.

Despite the discussed data limitations and differences, results of the EU model seem to point in the same direction as the Danish model for prioritizing interventions at the national farm-to-fork chain, showing almost the same order of importance for the sources common to both. The main difference was observed for turkeys, and it was not possible to evaluate which of the models present a more realistic estimation.

CONCLUSIONS

It was considered that results of the single-country model could be improved by the use of country-specific trade data for the *m* component, taking into consideration a weighted contribution of exporting countries to the number of cases attributed to the sources. The EU model would be improved by using more complete and representative data with a higher level of subtyping, but is still considered useful for countries which cannot readily attain the level of detailing found in Denmark for monitoring and surveillance data.

ACKNOWLEDGEMENTS

We would like to acknowledge the staff of EFSA's Task Force of Zoonoses Data Collection for providing the original datasets necessary to this study. The EU model was developed with funding from contract CT/EFSA/Zoonoses/2010/02 between EFSA and the DTU National Food Institute, in relation to Question n° EFSA-Q-2010-00685. We also acknowledge Timour Koupeev from Vose Risk Consulting for the collaboration on the management of the EUROSTAT data.

REFERENCES

Association of Poultry Processors and Poultry Trade in the EU Countries. Annual Report 2009. Brussels, Belgium: Association of Poultry Processors and Poultry Trade in the EU Countries.

Agersø Y, Hald T, Borck Høg B, Jensen LB, Jensen VF, Korsgaard H, Larsen LS, Pires SM, Seyfarth AM, Struve T, Hammerum AM, Jensen US, Lambertsen LM, Rhod Larsen A, Møller Nielsen E, Olsen SS, Petersen A, Skjøt-Rasmussen L, Skov RL, Sørum M. DANMAP 2010: DANMAP 2010 - Use of antimicrobial agents and occurrence of antimicrobial resistance in bacteria from food animals, food and humans in Denmark. National Food institute, 2011. 156 p.

Anonymous. Annual report on zoonoses in Denmark 1998. Copenhagen, Denmark: The Ministry of Food, Agriculture and Fisheries, 1999.

Anonymous. Annual report on zoonoses in Denmark 2010. Copenhagen, Denmark: The Ministry of Food, Agriculture and Fisheries, 2011.

de Jong B, Ekdahl K. The comparative burden of salmonellosis in the European Union member states, associated and candidate countries. BMC Public Health 2006; 6:4.

de Knegt LV. A multi-country approach for attributing human salmonellosis to animla reservoirs: global perspectives and application of surveillance data from the European Union (thesis). Kongens Lyngby, Denmark: Technical University of Denmark, 2013, 109p.

European Food Safety Authority. Report of the task force on zoonoses data collection on the analysis of the baseline survey on the prevalence of *Salmonella* in slaughter pigs in the EU, 2006-2007. Part A: *Salmonella* prevalence estimates. The EFSA Journal 2008a; 135:1-111.

European Food Safety Authority. Report of the task force on zoonoses data collection on the analysis of the baseline survey on the prevalence of *Salmonella* in turkey flocks, in the EU, 2006-2007. Part A: *Salmonella* prevalence estimates. EFSA Journal 2008b; 134:1-224.

European Food Safety Authority. The community summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in the European Union in 2008. EFSA Journal 2010b; 8:1496.

European Food Safety Authority. Analysis of the baseline survey on the prevalence of *Campylobacter* in broiler batches and of *Campylobacter* and *Salmonella* on broiler carcasses in the EU, 2008 - Part A: *Campylobacter* and *Salmonella* prevalence estimates. EFSA Journal 2010a; 8:1503.

Hald T, Vose D, Wegener HC, Koupeev T. A Bayesian approach to quantify the contribution of animal-food sources to human salmonellosis. Risk Anal. 2004; 24: 255-269.

Havelaar AH, Ivarsson S, Löfdahl M, Nauta M. Estimating the true incidence of campylobacteriosis and salmonellosis in the European Union, 2009. Epidemiol. Infect. 2012; 141: 293-302.

Pires SM, Hald T. Assessing the differences in public-health impact of *Salmonella* subtypes using a Bayesian microbial subtyping approach for source attribution. Food. Pat.Dis. 2010; 7: 143-151.

Pires SM., de Knegt LV, Hald T. 2011. Estimation of the relative contribution of different food and animal sources to human *Salmonella* infections in the European Union. Scientific Technical Report submitted to EFSA. EFSA-Q-2010-00685. Available at http://www.efsa.europa.eu/en/supporting/doc/184e.pdf.

Statistical Office of the European Union (EUROSTAT). (epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database). Acessed 25 March 2011.

Wegener H. Danish initiatives to improve the safety of meat products. Meat Science 2010; 84:276-283.

World Health Organization Collaborating Center for Reference and Research on *Salmonella*. Antigenic formulae of the *Salmonella* serovars. 9th Edition. Paris, France: Institut Pasteur, 2007.

Tables

Table 1. Data-related features of the Danish source account model and the EU source attribution model.

	Danish model	EU model	Comment
Human data included	Data from 2007, 2008 and 2009	Aggregated and case-based data from 2007 to 2009	No year-specific inferences are possible in the EU-model.
Source of human data	Statens Serum Institute (SSI).	ECDC / EFSA after reporting from countries.	Locally produced and reported data have fewer opportunities for information loss from the point of collection to the point of storage. Danish data in the EU model was reported to ECDC by SSI.
Travel information	Cases with missing information modeled according to the probabilities observed in the ones with information, resulting in 31% travel and 69% domestic.	Cases with missing information assumed to be domestic, resulting in 18% travel and 82% domestic.	The Danish model assumes that the follows the same distribution as the information provided. The EU model assumes that cases not referred specifically as travel-related are domestic, mainly because some countries had 0% travel information, and it was not possible to estimate the proportion of travellers. This assumption is likely to result in an underestimation of travel cases in the EU model, as some of the not-specified cases would be travel-related.
Subtyping information	Most isolates serotyped, *S*. Enteritidis and *S*. Typhimurium phage typed and *S*. Typhimurium tested for susceptibility to nine antimicrobials.	Serovar level used. The serovar distribution of cases and samples with missing serovar information were modeled based on observed distributions in the relevant datasets, resulting in a larger uncertainty regarding the true serovar distribution. This was particularly the case in the human datasets.	Higher level of detailing attributes cases more specifically to the right sources, but also leave a relatively higher proportion of cases with "unknown source", as the model requires a "perfect match" between subtypes in humans and animal reservoirs. On the other hand, in the model with less subtype detailing, cases could be misplaced, as the same serovar can be present in different sources, and the source with higher prevalence will ''draw" more cases.
Food/animal sources included and origin of *Salmonella* prevalence data,	Domestic: pork, beef, broilers, layers and ducks (from national surveillance programs). Imported: pork, beef, chicken, ducks and turkeys (from the case-by-case risk assessment program and retail monitoring).	Pigs, broilers, turkeys and layers (from EFSA baseline studies or EU-harmonized surveillance). Differentiation between imported and domestic based on the EUROSTAT production and trade data (see below).	The fewer the number of sources included in the model, the more likely it is for cases to be attributed to a wrong source. As an example, beef is absent from the EU model; however, *S*. Typhimurium is an important serovar in both cattle and pigs, and it is likely that some *S*. Typhimurium cases which were caused be beef are attributed to pigs in this model. Another expected resulting difference of the two approaches is that in the Danish modelimported eggs are not included, as theyare generally considered to be of low importance, as they are mainly used for heat-treated products by the industry and there is consequently no monitoring of imported shell eggs. O; in the EU model, they enter as a source, where the impact is determined by the imported amount and the prevalence in the country of origin.
Consumption data	Domestic and imported amounts of each source available in the country, with no differentiation between countries of origin of imported food.	Estimated from production, exports and imports reported to EUROSTAT. Specific amounts originating from each exporting country available.	The use of trade data, allows discrimination among foods originating from different countries, particularly when country-specific prevalences are available from the BS studies. The use of these data bring along some biases and assumptions, as described in the methods section.
Model dimensions	Subtype (serovar, phage type, resistance pattern), source and year	Serovar, source, country of human case reporting and country of origin of food	The "country of origin of food" dimension allows the attribution of cases to the country of origin of the sources.

Table 2. DK model: Estimated percentage of *Salmonella* cases attributed to food/animal sources, international travel, outbreaks with source unknown and unknown sources, 2007-2009, Denmark (mean and 95% Credibility Interval).

Source	2007	2008	2009	2007-2009
Broilers	0.3 (0.0-1.0)	1.3 (0.7-3.6)	0.7 (0.1-1.8)	0.9
Imported chicken	1.4 (0.4-2.8)	5.2 (3.3-6.8)	3.7 (2.1-5.3)	3.8
Pork	7.6 (6.0-9.3)	8.8 (7.6-10.0)	6.5 (3.6-9.7)	7.9
Imported pork	2.0 (1.0-3.1)	0.5 (0.3-1.9)	1.3 (0.2-2.8)	1.4
Turkeys	-	0	-	0
Imported turkey	2.0 (0.5-3.5)	2.4 (0.2-4.1)	0.7 (0.1-1.8)	1.9
Table eggs	12.3 (11.5-13.2)	3.2 (2.5-3.9)	11.0 (8.9-13.2)	7.5
Beef	0.2 (0.1-0.3)	0.7 (0.4-1.0)	0.7 (0.1-1.6)	0.6
Imported beef	3.1 (2.2-4.0)	0.3 (0.1-0.7)	1.2 (0.6-1.8)	1.3
Ducks	0.3 (0.0-0.9)	1.0 (0.1-2.7)		0.6
Imported duck	1.4 (0.5-2.3)			0.4
Travel	32.2 (30.4-31.4)	23.3 (23.1-23.6)	46.3 (44.4-48.2)	30.6
Unknown source	17.7 (15.1-19.8)	13.1 (11.3-15.0)	23.4 (20.0-26.8)	16.7
Outbreaks, unknown source	20.9	39.6	4.4	26.4
TOTAL	2,129	3,656	1,647	7,433

Table 3. EU model: Estimated percentage of *Salmonella* cases attributed to animal reservoirs, international travel, outbreaks with source unknown and unknown sources, 2007-2009, Denmark.

Source	Total source percentage[(a)]	Percentage by origin[(b)]
Broilers	3.5 (0.5-12,5)	0
Imported broilers		3.5
Pigs	18.0 (3.2-61.4)	14.7
Imported pigs		3.3
Turkeys	19.6 (2.9-69.2)	0
Imported turkeys		19.6
Layers	10.1 (2.3-33.1)	1.2
Imported eggs		8.9
Travel	23.7 (3.6-83.0)	23.7
Unknown source	18.3 (2.8-64.0)	18.3
Outbreaks, unknown source	6.8	6.8
Total	26,330	26,330

(a) Results of the EU model (de Knegt, 2003).

(b) Total source percentage divided based on country "originating" Danish cases (de Knegt, 2003).

.

Table 4. Comparison of the overall Salmonella prevalence and amount available for consumption in the two models and percentage of the number of cases reported in Denmark attributed to turkeys by the EU model

Exporting country	% of Danish cases attributed to turkeys	EU model		DK model	
		prevalence	*m*	prevalence	*m*
BE	0.1	10.8	80	N/A	N/A
DE	24.3	7.3	14,102	15.6	N/A
ES	1.5	39.1	222	N/A	N/A
FR	23.6	9.6	6,021	8.7	N/A
HU	7.2	62.5	782	N/A	N/A
IE	0.1	22.7	228	N/A	N/A
IT	9.3	20.2	2,968	46.4	N/A
LT	0.1	4.4	77	N/A	N/A
NL	0.9	9.0	512	N/A	N/A
PL	30.9	17.7	5,695	39.4	N/A
UK	1.9	25.5	783	N/A	N/A
Total	100.0	-	31,470	18.75	23,687

Table 5. Prevalence of selected serovars in the four animal sources included in the EU model in Poland (PL), Germany (DE) and France (FR)

Country	Serovar	Prevalence (p)			
		Broilers	Pigs	Turkeys	Layers
PL	*S.* Saintpaul	0.24	0.00	6.77	0.01
	S. Enteritidis	7.16	2.47	0.93	10.11
	S. Typhimurium	2.39	1.19	3.04	0.52
DE	*S.* Saintpaul	0.00	0.00	1.57	0.00
	S. Enteritidis	0.00	0.40	0.14	2.50
	S. Typhimurium	4.86	9.19	1.82	0.39
FR	*S.* Saintpaul	0.00	0.09	0.61	0.03
	S. Enteritidis	0.24	0.18	1.17	2.18
	S. Typhimurium	0.00	7.83	1.47	1.31

Table 6. Comparison of the overall Salmonella prevalence and amount available for consumption in the two models and percentage of the number of cases reported in Denmark attributed to broilers by the EU model

Exportin	% of Danish cases	EU model		DK model	
country	attributed to broilers	prevalence	m	prevalence	m
AR[(a)]	-	N/A	N/A	6.7	N/A
BE	9.4	20.3	7,335	8.4	N/A
BR[(a)]	-	N/A	N/A	12.0	N/A
CZ	0.4	5.5	416	N/A	N/A
DE	16.5	17.6	26,935	9.2	N/A
ES	7.2	14.9	2,418	N/A	N/A
FR	2.1	7.6	8,644	10.0	N/A
GR	0.1	14.8	70	N/A	N/A
HU	9.7	85.7	1,442	N/A	N/A
IE	2.0	9.9	218	N/A	N/A
IT	0.4	16.8	894	N/A	N/A
LT	0.5	6.9	2,299	6.6	N/A
LV	0.1	4.9	27	N/A	N/A
NL	2.6	10.0	23,773	42.9	N/A
PL	30.2	25.5	6,597	3.6	N/A
PT	7.0	11.2	1,633	N/A	N/A
SE	0.4	0.2	71,499	4.9	N/A
SI	1.0	1.7	3,426	N/A	N/A
SK	0.3	21.6	51	N/A	N/A
UK	9.9	3.5	8,287	N/A	N/A
Total	100.0	-	165,964	8.6	93,191

(a) Non-EU countries from where Denmark has imported chicken meat

Table 7. Comparison of the overall Salmonella prevalence and amount available for consumption in the two models and percentage of the number of cases reported in Denmark attributed to pigs by the EU model

Exporting country	% of Danish cases attributed to pigs	EU model		DK model	
		prevalence	m	prevalence	m
BE	0.6	13.0	11,840	N/A	N/A
DE	6.8	12.7	123,767	10.0	N/A
DK	81.5	8.0	3,013,472	3.1	N/A
ES	5.9	30.7	62,648	33.3	N/A
FR	1.1	18.5	22,896	29.6	N/A
HU	0.3	11.6	3,611	N/A	N/A
IE	0.6	15.4	10,592	N/A	N/A
IT	0.1	16.4	4,355	N/A	N/A
NL	1.4	8.5	46,638	16.7	N/A
PL	0.7	0.7	11,069	N/A	N/A
UK	1.0	1.0	12,969	31.8	N/A
Total	100.0	-	3,323,857	11.9	230,440

Table 8. Comparison of the overall Salmonella prevalence and amount available for consumption in the two models and percentage of the number of cases reported in Denmark attributed to layers by the EU model

Exporting country	% of cases Danish cases attributed to layers	EU model		DK model	
		prevalence	m	prevalence	m
AT	0.1	2.5	341	N/A	N/A
BE	0.9	11.7	1,060	N/A	N/A
CZ	0.3	8.9	167	N/A	N/A
DE	4.8	3.5	8,999	N/A	N/A
DK	11.9	0.6	200,645	5.42	82,594
ES	5.8	44.5	1,080	N/A	N/A
FR	0.1	6.1	121	N/A	N/A
LV	2.3	20.3	829	N/A	N/A
NL	4.1	2.6	7,595	N/A	N/A
PL	69.6	12.5	32,450	N/A	N/A
SE	0.1	0.7	2,763	N/A	N/A
Total	100.0	-	256,050	N/A	82,594

Figures

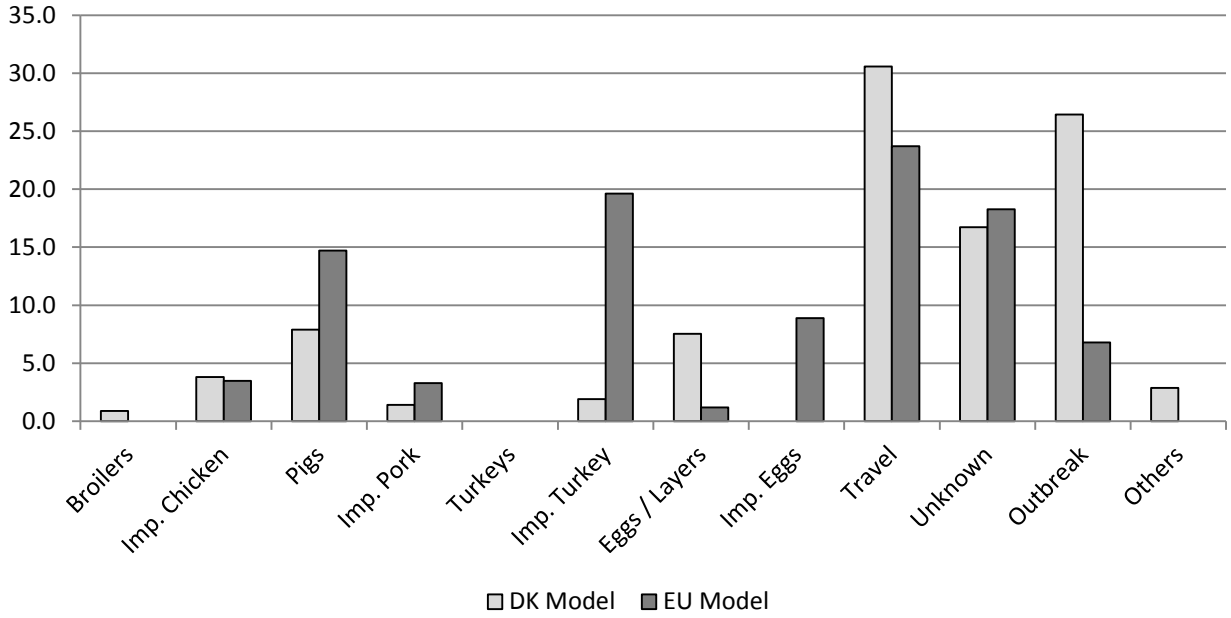


Figure 1. Attributable fractions of *Salmonella* cases to domestic and imported animal sources in Denmark in the Danish model and in the EU model, 2007-2009.

APPENDICES

Appendix A. Country-specific attribution estimates to food-animal reservoirs, travel, outbreaks and unknown sources

Appendix B. Attribution estimates to food-animal reservoirs, travel, outbreaks and unknown sources in the EU regions

Appendix C. Proportion of cases attributed to the main serovars in broilers, pigs, turkeys and layers

Appendix D. Attribution estimates to food-animal reservoirs in their country of origin. The percentage column refers to percentage of EU cases "originated" by that country

Appendix E. Percentage of cases attributed to each source in the EU "originating" from each country

Appendix F. Estimated percentage of sporadic Salmonella cases in Denmark attributed to animal reservoirs originating from exporting countries and Denmark, 2007-2009

Appendix G. Elicitation instrument

Appendix H. Sources and description of the data used for the cluster analyses

Appendix I. Demonstration sheets provided along with the information sheets containing results of the cluster analyses

Appendix J. Elicited estimates for Bulgaria, Norway and Romania

Appendix K. Individual expert guesses (most likely value, minimum and maximum) in Bulgaria (BG1 to BG4), Norway (NO1 to NO4) and Romania (RO1 to RO4)

Appendix L. Joint panel estimate distributions in Bulgaria (L1), Norway (L2) and Romania (L3)

Appendix A. Country-specific attribution estimates to food-animal reservoirs, travel, outbreaks and unknown sources.

	AT				BE				CY			
Source	**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**	
Broilers	0.1	73	26	277	2.3	935	104	3,672	4.8	4,226	544	15,500
Pigs	14.4	13,130	1,971	45,970	74.2	30,130	3,461	117,300	51.1	44,580	6,639	156,700
Turkeys	3.7	3,417	503	12,090	9.2	3,750	423	14,680	6.4	5,626	618	21,480
Layers	59.8	54,520	8,310	189,500	2.9	1,178	123	4,710	8.9	7,722	976	28,520
Travel	12.2	11,110	1,674	38,690	0.0	0	0	0	3.8	3,334	504	11,650
Unknown	9.4	8,605	1,267	30,210	11.2	4,554	512	17,810	24.9	21,750	3,128	77,430
Outbreak	0.3	272			0.1	52			0.0	0		

	CZ				DE				DK			
Source	**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**	
Broilers	0.1	1,308	92	5,201	0.5	6,378	519	24,980	3.5	918	132	3,295
Pigs	10.9	128,900	19,490	446,700	33.1	420,300	63,750	1,462,000	18.0	4,743	854	16,170
Turkeys	1.8	20,710	3,080	72,250	1.3	17,000	2,561	59,330	19.6	5,167	775	18,210
Layers	84.6	997,000	151,300	3,450,000	52.0	660,800	100,100	2,301,000	10.1	2,665	617	8,710
Travel	1.7	20,090	3,047	69,610	5.3	67,860	10,260	236,100	23.7	6,239	946	21,850
Unknown	0.8	9,890	-1,204	41,970	7.6	96,850	14,570	337,000	18.3	4,813	725	16,860
Outbreak	0.0	88			0.2	1,990			6.8	1,786		

	EE				ES				FI			
Source	**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**	
Broilers	4.6	923	160	3,356	0.1	3,384	45	17,680	0.7	21	0	96
Pigs	27.5	5,488	818	19,130	33.1	869,600	130,000	3,066,000	4.7	150	22	530
Turkeys	2.1	421	47	1,601	12.9	339,100	50,400	1,196,000	1.6	53	5	203
Layers	55.0	10,980	1,671	37,940	43.1	1,133,000	169,200	4,003,000	2.4	79	10	291
Travel	7.9	1,587	244	5,460	0.0	0	0	0	80.1	2,571	387	8,939
Unknown	2.6	516	-601	2,764	10.7	281,100	41,470	993,700	4.6	148	21	530
Outbreak	0.3	63			0.0	469			5.9	189		

	FR				GR				HU			
Source	**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**	
Broilers	13.4	66,000	10,120	230,000	1.2	28,530	384	148,100	4.5	52,570	7,904	182,900
Pigs	34.3	168,900	25,950	586,700	9.5	227,200	33,520	801,600	26.7	313,300	47,160	1,090,000
Turkeys	12.6	62,180	9,363	217,400	0.4	9,061	468	40,570	5.4	63,760	9,558	222,200
Layers	2.9	14,150	2,864	47,600	78.3	1,872,000	279,200	6,552,000	54.9	643,600	96,960	2,231,000
Travel	0.0	0	0	0	2.3	55,820	8,336	195,400	0.2	1,975	298	6,840
Unknown	36.5	179,800	27,140	627,000	8.3	197,700	25,090	721,300	8.1	94,870	14,110	331,500
Outbreak	0.2	966			0.0	0			0.2	1,815		

	IE				IT				LT			
Source	**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**	
Broilers	1.5	100	6	486	2.3	17,680	2,639	62,030	1.2	5,244	631	19,630
Pigs	27.2	1,810	113	8,616	73.2	560,700	85,200	1,949,000	9.5	42,750	6,428	148,700
Turkeys	8.8	589	35	2,810	5.3	40,410	6,028	141,700	0.7	3,318	369	12,600
Layers	14.6	971	64	4,594	2.2	16,520	2,309	59,450	86.9	390,000	59,010	1,353,000
Travel	31.7	2,110	133	10,020	1.3	9,908	1,505	34,480	0.3	1,294	196	4,488
Unknown	15.3	1,018	62	4,864	15.8	120,800	18,280	421,300	1.2	5,596	-4,449	27,910
Outbreak	0.9	63			0.0	0			0.1	335		

	LU				LV				NL			
Source	**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**	
Broilers	4.4	96	6	449	0.9	873	92	4,135	4.6	4,455	653	15,810
Pigs	8.5	184	13	833	13.7	13,590	2,052	47,280	27.3	26,330	3,978	91,590
Turkeys	6.9	149	11	670	0.3	291	6	1,368	9.7	9,404	1,393	33,050
Layers	49.8	1,073	89	4,662	82.5	81,600	12,450	282,200	26.2	25,270	4,015	87,770
Travel	9.6	207	17	896	1.5	1,459	222	5,046	14.2	13,730	2,079	47,900
Unknown	20.7	446	35	1,961	0.7	714	-4,236	7,337	17.5	16,920	2,521	59,240
Outbreak	0.0	0			0.4	351			0.5	470		

	PL				PT				SE			
Source	**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**	
Broilers	25.1	796,600	120,900	2,772,000	42.3	1,357,000	202,700	4,727,000	0.5	28	2	117
Pigs	47.8	1,520,000	229,700	5,269,000	36.3	1,164,000	175,500	4,052,000	4.8	282	42	991
Turkeys	1.2	39,640	5,790	139,900	0.6	18,580	546	83,890	1.7	99	13	361
Layers	23.0	731,300	111,500	2,550,000	9.1	290,400	29,270	1,138,000	2.5	145	29	506
Travel	0.1	1,978	300	6,882	0.4	11,250	1,704	39,030	75.9	4,441	666	15,530
Unknown	2.7	84,840	11,520	305,300	11.4	364,300	49,970	1,310,000	10.2	596	89	2,089
Outbreak	0.1	3,484			0.0	90			4.4	260		

	SI				SK				UK			
Source	**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**	
Broilers	0.5	564	7	3,037	0.0	363	21	1,779	0.6	1,590	236	5,565
Pigs	20.6	21,600	2,464	84,410	18.0	189,300	28,490	664,900	11.7	32,370	4,886	112,600
Turkeys	4.0	4,197	452	16,740	2.6	27,580	4,066	97,930	10.1	27,930	4,208	97,290
Layers	59.5	62,240	7,195	242,500	76.8	807,500	121,800	2,826,000	35.5	97,990	14,800	340,900
Travel	0.0	0	0	0	0.8	8,152	1,228	28,540	24.3	67,250	10,170	234,100
Unknown	14.7	15,370	1,716	60,450	1.7	17,940	1,124	70,500	17.8	49,270	7,449	171,200
Outbreak	0.6	656			0.0	449			0.0	0		

Appendix B. Attribution estimates to food-animal reservoirs, travel, outbreaks[(a)] and unknown sources in the EU regions[(b)].

	Eastern EU				Northern EU			
	%	mean	95% CI		%	mean	95% CI	
Broilers	12.9	850,800	162,700	2,828,000	1.1	9,696	3,363	24,930
Pigs	32.7	2,152,000	657,800	5,993,000	11.4	101,200	39,130	235,400
Turkey	2.3	151,700	56,120	350,600	4.3	37,870	10,970	108,300
Layers	48.3	3,179,000	1,234,000	6,994,000	66.0	584,500	181,900	1,575,000
Outbreaks	0.1	5,836			0.3	3,047		
Travel	0.5	32,200	10,160	84,850	9.8	86,950	26,120	254,100
Unknown	3.2	207,500	65,260	532,200	7.1	62,670	15,300	186,300

	Western EU				Southern EU			
	%	mean	95% CI		%	mean	95% CI	
Broilers	3.9	77,930	18,930	242,600	15.4	1,411,000	249,000	4,783,000
Pigs	33.1	658,900	207,100	1,761,000	31.5	2,888,000	1,089,000	6,655,000
Turkey	4.8	95,900	31,710	256,000	4.5	416,900	108,000	1,279,000
Layers	38.0	757,000	178,200	2,397,000	36.8	3,382,000	1,054,000	8,832,000
Outbreaks	0.2	3,750			0.0	1,215		
Travel	4.7	92,900	26,390	263,400	0.9	80,310	24,620	222,400
Unknown	15.4	307,200	97,720	809,100	10.9	1,001,000	384,100	2,277,000

(a) The proportion of outbreak cases were derived directly from the reported data (i.e. they were not estimated and consequently no Credibility Intervals were calculated); includes outbreaks with unknown source. Outbreak cases for which the source was identified were assigned to the correspondent animal sources.

(b) EU regions as defined by the United Nations. Eastern Europe: Czech Republic, Hungary, Poland and Slovakia. Northern Europe: Denmark, Estonia, Finland, Ireland, Latvia, Lithuania, Sweden and the United Kingdom. Southern Europe: Cyprus, Greece, Italy, Portugal, Slovenia, Spain. Western Europe: Austria, Belgium, France, Germany, Luxembourg and the Netherlands.

Appendix C. Proportion of cases attributed to the main serovars in broilers, pigs, turkeys and layers.

Broilers		Pigs		Turkeys		Layers	
Serovar	**%**	**Serovar**	**%**	**Serovar**	**%**	**Serovar**	**%**
***S.* Enteritidis**	85.0	***S.* Typhimurium**	50.9	***S.* Enteritidis**	27.9	***S.* Enteritidis**	95.0
***S.* Infantis**	4.5	***S.* Enteritidis**	38.2	***S.* Typhimurium**	18.6	***S.* Typhimurium**	1.4
***S.* Typhimurium**	2.5	***S.* Derby**	1.8	***S.* Newport**	9.2	***S.* Infantis**	1.3
***S.* Virchow**	2.9	***S.* Infantis**	1.1	***S.* Saintpaul**	4.5	***S.* Virchow**	1.0
***S.* Kentucky**	0.6	***S.* Newport**	2.3	***S.* Hadar**	19.0	***S.* Kentucky**	0.2
Others	4.5	**Others**	5.7	**Others**	21.0	**Others**	1.0
Total cases	2,350,000	**Total cases**	5,800,000	**Total cases**	702,400	**Total cases**	7,903,000

Appendix D. Attribution estimates to food-animal reservoirs in their country of origin. The percentage column refers to percentage of EU cases "originated" by that country.

	AT				BE				CY			
Source	**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**	
Broilers	0.3	191	67	466	0.8	16,540	5,603	44,550	0.3	4,155	552	15,190
Pigs		23,560	10,910	47,650		109,000	53,050	220,300		40,090	10,190	122,900
Turkeys		2,437	947	5,810		851	258	2,516		456	50	1,741
Layers		31,970	6,051	107,300		14,340	3,923	43,140		3,045	386	11,220

	CZ				DE				DK			
Source	**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**	
Broilers	6.0	956	247	3,045	6.7	7,650	2,587	20,900	0.5	0	0	0
Pigs		114,700	29,510	323,100		645,100	265,700	1,551,000		85,460	37,260	189,400
Turkeys		15,020	4,451	40,000		22,310	10,170	50,260		0	0	0
Layers		874,200	142,000	2,999,000		440,100	124,900	1,258,000		584	226	1,390

	EE				ES				FI			
Source	**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**	
Broilers	0.1	0	0	0	18.4	58,490	13,020	185,800	0.0	0	0	0
Pigs		9,121	3,271	21,790		1,306,000	423,700	3,556,000		0	0	0
Turkeys		0	0	0		302,600	55,350	1,029,000		0	0	0
Layers		5,419	1,339	16,020		1,414,000	406,100	4,286,000		10	4	22

	FR			GR			HU		
Source	**%**	**mean**	**95% CI**	**%**	**mean**	**95% CI**	**%**	**mean**	**95% CI**
Broilers	2.5	35,210	6,162 120,300	10.9	27,380	499 141,600	6.0	53,660	9,881 180,800
Pigs		238,400	91,980 576,800		90,550	13,560 319,100		286,600	59,800 943,700
Turkeys		116,700	43,460 287,300		445	54 1,754		84,060	27,580 230,500
Layers		20,610	7,262 52,790		1,701,000	256,400 5,944,000		587,900	93,970 2,023,000

	IE			IT			LT		
Source	**%**	**mean**	**95% CI**	**%**	**mean**	**95% CI**	**%**	**mean**	**95% CI**
Broilers	0.1	3,927	1,474 9,996	2.4	17,440	3,194 59,030	0.0	1,192	216 4,185
Pigs		8,004	4,158 15,200		299,900	51,940 1,024,000		4,684	791 16,020
Turkeys		638	185 1,809		56,860	19,810 153,800		399	108 1,207
Layers		7	1 21		32,510	10,850 82,980		0	0 0

	LU			LV			NL		
Source	**%**	**mean**	**95% CI**	**%**	**mean**	**95% CI**	**%**	**mean**	**95% CI**
Broilers	0.0	0	0 0	1.2	876	157 3,126	1.8	1,944	890 4,067
Pigs		340	146 785		3,016	544 10,190		121,000	56,900 251,200
Turkeys		0	0 0		0	0 0		5,088	2,711 9,397
Layers		414	39 1,776		196,500	69,420 486,500		165,200	39,940 512,100

	PL				PT				SE			
Source	**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**	
Broilers	21.3	803,600	131,400	2,768,000	14.5	1,305,000	198,500	4,535,000	0.0	7	2	20
Pigs		1,402,000	257,400	4,721,000		876,000	134,800	3,040,000		364	189	695
Turkeys		71,110	30,950	167,900		1,342	198	5,397		0	0	0
Layers		1,287,000	492,000	3,162,000		239,800	27,870	928,000		64	13	215

	SI				SK				UK			
Source	**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**		**%**	**mean**	**95% CI**	
Broilers	0.4	412	33	2,060	5.0	1,711	628	3,999	0.8	8,632	3,920	18,300
Pigs		11,440	1,577	43,820		72,300	12,190	249,600		50,810	20,800	117,600
Turkeys		2,864	381	11,170		71	18	220		19,080	6,737	52,040
Layers		57,020	6,929	221,100		768,300	192,800	2,339,000		60,270	10,210	206,300

Appendix E. Percentage of cases attributed to each source in the EU "originating" from each country.

Country	Percentage of cases attributed to source			
	Broilers	**Pigs**	**Turkey**	**Layers**
AT	0.0	0.4	0.3	0.4
BE	0.7	1.9	0.1	0.2
CY	0.2	0.7	0.1	0.0
CZ	0.0	2.0	2.1	11.1
DE	0.3	11.1	3.2	5.6
DK	0.0	1.5	0.0	0.0
EE	0.0	0.2	0.0	0.1
ES	2.5	22.5	43.1	17.9
FI	0.0	0.0	0.0	0.0
FR	1.5	4.1	16.6	0.3
GR	1.2	1.6	0.1	21.5
HU	2.3	4.9	12.0	7.4
IE	0.2	0.1	0.1	0.0
IT	0.7	5.2	8.1	0.4
LT	0.1	0.1	0.1	0.0
LU	0.0	0.0	0.0	0.0
LV	0.0	0.1	0.0	2.5
NL	0.1	2.1	0.7	2.1
PL	34.2	24.2	10.1	16.3
PT	55.6	15.1	0.2	3.0
SE	0.0	0.0	0.0	0.0
SI	0.0	0.2	0.4	0.7
SK	0.1	1.2	0.0	9.7
UK	0.4	0.9	2.7	0.8
Total	100.0	100.0	100.0	100.0

Appendix F. Estimated percentage of sporadic *Salmonella* cases in Denmark attributed to animal reservoirs originating from exporting countries and Denmark, 2007-2009.

Exporting country	Broilers	Pigs	Turkeys	Layers
Austria	0.0	0.0	0.0	0.1
Belgium	9.4	0.6	0.1	0.9
Cyprus	0.0	0.0	0.0	0.0
Czech Republic	0.4	0.0	0.0	0.3
Germany	16.5	6.8	24.3	4.8
Denmark	0.0	81.5	0.0	11.9
Estonia	0.0	0.0	0.0	0.0
Spain	7.2	5.9	1.5	5.8
Finland	0.0	0.0	0.0	0.0
France	2.1	1.1	23.6	0.1
Greece	0.1	0.0	0.0	0.0
Hungary	9.7	0.3	7.2	0.0
Ireland	2.0	0.6	0.1	0.0
Italy	0.4	0.1	9.3	0.0
Lithuania	0.5	0.0	0.1	0.0
Luxembourg	0.0	0.0	0.0	0.0
Latvia	0.1	0.0	0.0	2.3
Netherlands	2.6	1.4	0.9	4.1
Poland	30.2	0.7	30.9	69.6
Portugal	7.0	0.0	0.0	0.0
Sweden	0.4	0.0	0.0	0.1
Slovenia	1.0	0.0	0.0	0.0
Slovakia	0.3	0.0	0.0	0.0
United Kingdom	9.9	1.0	1.9	0.0
Total	100.0	100.0	100.0	100.0

Appendix G. Elicitation instrument

Attribution of *Salmonella* cases in humans to animal reservoirs using expert elicitation with cluster analysis as an information tool

Dear expert,

You are being invited to participate in an expert elicitation to attribute human cases of salmonellosis to animal reservoirs of the food chain. Before proceeding to the questions, please read carefully the study description. If you have any questions regarding data origin or units, section 4 contains a spreadsheet with a variable dictionary.

Objectives:

- test and validate an approach in which easily-available data can be used to relate countries with no attribution results to countries for which traditional attribution studies have been performed, and obtain estimates for the first group through expert elicitation;
- evaluate whether this approach is useful and whether it provides sensible results

Section 1 – Approach and instructions

The information sheets received along with this document contain results of cluster analysis with information about 29 European countries in different combinations, depending on data availability. The sheets are divided in nine main groups:

1. Source attribution outcomes for 24 European Union Member States (Pires et al., 2011).
2. Relative proportions of S. Enteritids, S. Typhimurium and "Other serovars" in humans and animal sources in each country (Pires et al., 2011).
3. Food consumption information (FAO, 2003).
4. Economic indicators (UNDP, 2011).
5. Non-economic human development indicators (UNDP, 2011).
6. Agriculture and land usage characteristics (FAO, 2011).
7. Density of animal production (FAO, 2011).
8. Climate data. Peel et al. (2007).
9. Cluster results summary (Excel file).

Each sheet from groups 1 to 8 contains a table with the data used for the analysis and the division based on the best cluster solution. In multi-variable sheets, the variables that drove the formation of each cluster are highlighted. For easier visualization of the distance between countries not in the same cluster, dendrograms are also provided. The Excel file contains matrices summarizing how frequently each two countries were clustered together, as well as in which sheets that occurred. Experts will look at how countries relate to each other in the different sheets and in the matrices and:

1- provide attribution point-estimates for the selected sources in Bulgaria, the Czech Republic, Norway and Romania;
2- provide minimum and maximum expected values for the point estimates;
3- evaluate the information provided for the elicitation;
4- evaluate the usefulness of the method.

A more detailed description of the study background, methods and data origin can be found in sections 3 and 4.

Section 2 - Questions

1- Please, check the field you would describe as your main area of expertise:

a) Epidemiology / Risk modelling / Risk assessment
b) Infectious diseases / Microbiology / Parasitology
c) Toxicology / Food chemistry
d) Policy-making

2 - Fill in the attributable fractions (%) you estimate for each source in Bulgaria, the Czech Republic, Norway and Romania, adding a range for your answer. If not enough information was provided for an estimate, write "NP" (for "not possible") in the corresponding field.

Source	Czech Republic			Bulgaria		
	%	Range		%	Range	
Broilers						
Pigs						
Turkeys						
Layers						
Travel						
Unknown / other reservoirs						

Source	Romania			Norway		
	%	Range		%	Interval range	
Broilers						
Pigs						
Turkeys						
Layers						
Travel						
Unknown / other reservoirs						

2- Please, list at least one information sheet that:

a) was crucial for your decision (excluding the attribution results)

b) was crucial for your decision but you don't expect that information to be available for developing countries

c) did not contribute at all for your decision

d) provided information that you consider wrong or misleading

3- Do you consider this a valid approach for source attribution?

Section 3 – Study description

Background

During the last ten years, source attribution methods have become an important tool to provide risk assessors and managers with information for priority-targeting and policy-making (Havelaar et al. 2007; Kuchenmüller et al. 2009). As a consequence, different approaches have been developed for that purpose, such as microbial subtyping, analysis of outbreak data or case-control studies (Pires et al. 2009).

Among the most widespread statistical methods for source attribution of *Salmonella* are the Hald model and its variations. This method is based on microbial subtyping and has been adapted and applied in Denmark (Hald et al 2004; Pires and Hald 2010), Japan (Toyofuku et al. 2011), New Zealand (Mulner et al. 2009), Sweden (Whalström et al 2011) and the United States (Guo et al. 2011). A model including 24 European countries has been recently published (Pires et al 2011) as a technical report prepared for the European Food Safety Authority (EFSA).

One characteristic of the aforementioned models is that they require a large amount of good-quality data, which are available from the Danish surveillance system and, up to a point, from datasets maintained by Eurostat, studies published by EFSA and national harmonized surveillance systems. These data requirements reduce the applicability of such models in poor countries, where representative incidence or prevalence data is not so readily-available, and where the share of the burden of foodborne diseases is presumably larger than in Europe (Kuchenmüller et al. 2009).

When data required for a statistic approach seem to lack quality or is unavailable, expert elicitations can be used to obtain valid attribution estimates (Batz et al., 2005), and that has been done in several countries, such as the United States (Hoffmann et al, 2007), the Netherlands (Havelaar et al 2008), New Zealand (Lake et. al 2010) and Canada (Davidson et al, 2011). This study is an attempt to validate an alternative approach for source attribution of *Salmonella* that could be applied in situations where the data normally required for the traditional models cannot be obtained.

Cluster analysis is a technique used to group observations according to values observed for one or more variables. Although several methods and approaches can be used for that procedure, all of them are based on calculating the distance between each two observations and grouping the nearest ones according to a set of criteria. The procedure is then repeated using the distance between observations of two different clusters and regrouping them as one, and so on until on the bottom side of the tree each independent observation constitutes one small cluster, and on the top all observations are grouped as a large one, as described in Sharma (1996).

Our objective is to use easily-available data as a tool to fit countries with no attribution results into a general profile together with countries for which traditional attribution studies have been performed, and obtain estimates for the first group through expert elicitation. For the present exercise, source attribution estimates will be obtained for four European countries with different amounts and types of information available, as an extrapolation of the results observed in Pires et al. (2011). The point of attribution chosen was the animal reservoir, as a consequence of the method used and in line with the original study.

Methods

Microbial subtyping approach

The subtyping approach has so far been primarily applied to attribute foodborne pathogens. The approach involves characterization of the agent by subtyping methods (e.g., phenotypic or genotypic subtyping of bacterial pathogens), and the principle is to compare the subtypes of isolates from different sources (e.g., animals, food) with the same subtypes isolated from humans. The subtyping approach is enabled by the identification of strong associations between some of the dominant subtypes and a specific reservoir or source, providing a heterogeneous distribution of subtypes among the sources. Subtypes exclusively or almost exclusively isolated from one source are regarded as indicators for the human health impact of that particular source, assuming that all human cases caused by these subtypes originate only from that source. Human cases of disease caused by subtypes found in several reservoirs are then distributed relative to the prevalence of the indicator types. This approach utilizes a collection of temporally and spatially related isolates from various sources, and thus it is facilitated by an integrated foodborne disease surveillance programme that is focused on the collection of isolates from the major food animal reservoirs of foodborne diseases.

Approach assumption: illnesses caused by subtypes found exclusively in one source all originate from that source. The method therefore assumes that these subtypes do not occur in any other potential source of exposure (e.g. foods or other sources not sampled e.g. environment and wild animals) because no evidence has been reported.

Model applicability: the approach can only be applied to hazards that have multiple subtypes (e.g. bacteria phenotypic or genotypic subtypes) that are heterogeneously distributed among the sources.

Data requirements: the subtyping approach requires a representative distribution of the subtypes of the hazard in the potential sources and humans. It therefore relies on the existence of public health and animal/ environment surveillance, providing representative data for the covered region. Additionally, it is necessary that the same subtyping methods are applied to both human and source/foods isolates. The approach does not require estimates of the prevalence of the subtypes in each source.

The results used as reference to estimate source attribution parcels are originally from Pires et al. (2011), where the subtyping approach was used for attribution of *Salmonella* cases to broilers, pigs, turkeys and laying hens in 24 countries of the European Union. Cases that cannot be attributed to any of the reservoirs or have not been reported as travel-related are presented as "Unknown / Other sources". This also includes cases from outbreaks for which the source was not identified. Outbreaks for which the source has been identified have their cases added to the appropriate reservoir or to the previous group, if it was caused by some other source.

Salmonella prevalence in broilers, turkeys and pigs were obtained from the baseline studies published by the European Food Safety Authority (EFSA 2007; EFSA 2008b; EFSA 2010b). Human incidences were provided by the European Centers for Disease Control and Prevention through EFSA. Total human cases and *Salmonella* prevalence in layers are used as reported by countries to the Community summary Report on Trends and Zoonoses from 2007 to 2009, also published by EFSA (EFSA 2009: EFSA 2010a; EFSA 2011a). It is important to note here that the loss of data at various points along the surveillance chain from patient to official statistics is recognized in all countries (Wheeler et al., 1999) and results in different degrees of underreporting. This underreporting was compensated where necessary with the use of underreporting factors (EFSA, 2011b).

Cluster analysis

Hierarchical cluster analysis starts with all observations in a dataset belonging to the same cluster, and systematically creates new clusters, by separating observations which are more similar among themselves than to the remaining group in relation to a set of variables. The procedure is repeated until each observation constitutes its own cluster.

In this study, an "average subject" from each cluster was chosen as the centroid to be compared with other clusters, and the squared Euclidean distance between observations within the same cluster was used as similarity measure; the more similar the subjects, the smaller the distance between them (and consequently, the smaller the squared Euclidean distance) and vice-versa. Variables measured in different scales which were used in the same set were standardized to fit a distribution with mean=0 and standard deviation=1. It is necessary to standardize the values before running the analysis, otherwise variables that differ thousands of units from each other (e.g., country territory in squared kilometers) will drive the cluster construction, annulling the influence of variables that vary in a smaller scale (e.g., percentages).

The resulting process can be plotted as a dendrogram (or "tree") with the distance between clusters on the vertical axis. Although the whole hierarchical structure can be visualized in this way, the best cluster solution was chosen for each set of variables to be presented. This choice was based on an evaluation of the clustering process using a) the root-mean-square deviation (RMSSTD) of each new cluster formed, b) the semipartial R-squared (SPR), c) the R-squared (RS) and the distance between two clusters (CD). These measures provide a statistical reference to evaluate the homogeneity of a new cluster formed and the heterogeneity among the current group of clusters in each step, indicating the more "natural" number of clusters for a given set of observations.

The information sheets received along with this document contain results of cluster analyses with information about 29 European countries in different combinations, depending on data availability. They can be classified in nine major groups, which contain sheets with resulting clusters for one variable, as well as multi-variable sheets, which mean to provide a general profile of the countries concerning the way several factors combine. The main groups are:

1. *Source attribution outcomes. This group contains results of the source attribution approach based on microbial subtyping in 24 EU countries published in Pires et al. (2011), and should be used as a reference to estimate attributable fractions to animal reservoirs in countries without attribution studies. The incidences refer to a period of three years (2007-2009), and are given in cases/100000. Sheets included are:*
    a. *Salmonella* incidence attributable to all sources (overview table);
    b. *Salmonella* incidence attributable to broilers (table + dendrogram);
    c. *Salmonella* incidence attributable to pigs (table + dendrogram);
    d. *Salmonella* incidence attributable to turkeys (table + dendrogram);
    e. *Salmonella* incidence attributable to layers (table + dendrogram);
    f. attributable fraction of human *Salmonella* cases to all sources combined (overview table);
    g. cumulative attributable fractions bar graph;

Although the percentage of travel-related cases is shown in the bar graph, it was not included in the cluster analysis, as differences between countries were too large and obscured the importance of the contribution of animal reservoirs.

2. *Relative proportions of S. Enteritids, S. Typhimurium and "Other serovars" in humans and animal sources in each country* (Pires et al., 2011) *(5 sheets):*
    a. relative proportion of reported *S*. Enteritidis, *S*. Typhimurium and "Other serovars" in humans;
    b. relative proportion of *S*. Enteritidis, *S*. Typhimurium and "Other serovars" in broilers;
    c. relative proportion of *S*. Enteritidis, *S*. Typhimurium and "Other serovars" in pigs;
    d. relative proportion of *S*. Enteritidis, *S*. Typhimurium and "Other serovars" in turkeys;
    e. relative proportion of *S*. Enteritidis, *S*. Typhimurium and "Other serovars" in layers;

3. *Food consumption information* (FAO, 2003) *(4 sheets)*:
    a. relative proportions of consumption of eggs, poultry meat, pork, beef, sheep & goat, fish, seafood, raw animal fats and "other meats" ;
    b. consumption of poultry meat (g/person/day);
    c. consumption of pork (g/person/day);
    d. consumption of eggs (g/person/day);

4. *Economic indicators* (UNDP, 2011). *This group contains one sheet in which countries were clustered according to three variables*:
    a. gross domestic product (GDP) per capita in U.S. dollars;
    b. percentage of the population which is economically active;
    c. percentage of the population below the national poverty line;

5. *Non-economic human development indicators* (UNDP, 2011). *This group contains one sheet in which countries were clustered according to four variables*:
    a. literacy rate (%);
    b. mean years of schooling among adults;
    c. life expectancy in years;
    d. mortality under five years of age (per 1000 births);

6. *Agriculture and land usage characteristics* (FAO, 2011). *This group contains one tree in which countries were clustered according to four variables*:
    a. percentage of country territory used for agriculture;
    b. percentage of economically active population working full-time in agriculture;
    c. number of farms per square kilometer of agricultural land;
    d. number of individuals employed full time in agriculture per farm unit;

7. *Density of animal production* (FAO, 2011). *This group contains one sheet in which countries were clustered according to three variables together*:
    a. chickens per farm;
    b. pigs per farm;
    c. turkeys per farm;

8. *Climate data*. *This sheet contains a map of Europe showing Köppen-Geiger climate zones as updated by Peel et al. (2007), as well as a table extracted from the original article with a description of Köppen climate symbols and defining criteria. No cluster analysis was performed, as national borders and climate zones do not always coincide.*
9. *Cluster results summary (Excel file). This group contains "country X country" matrices based on the best solution for each set of variables, showing:*
    a. in which information sheets each two countries belonged in the same cluster;
    b. the probability that two countries belonged in the same cluster in the study, calculated by dividing the number of times they were clustered by the number of times they could be clustered, as not every country was present in every analysis.

Numbers in the gray cells refer to when a country formed its own individual cluster. Climate data was not included in the matrices.

Appendix H. Sources and description of the data used for the cluster analyses

Variable	Description	Unit	Obs
Country	Country	N/A	
Population	Population	Persons	WHO, 2011
Attributable fractions	Proportion of human cases attributed to each source in the EU Attribution model	%	EFSA-Q-2010-00685. Available at (http://www.efsa.europa.eu/en/supporting/pub/184e.htm?WT.mc_id=EFSAHL01&emt=1)
Attributable incidence	Incidence of cases attributed to each animal reservoir	Cases/100,000 pop	Attributable fraction applied to the number of cases reported in the CSR 2007-2009 and divided by the population *100,000, adjusted with underreporting factors from Havelaar et al. (2012)
Source consumption	Estimated amount of each source consumed in the country.	g/person/day	FAO, 2003. Broilers and turkeys together as Poultry
GDP Per Capita	Gross Domestic Product per capita	Euro/person	Eurostat 2011 - obs 2010 figures.
Percentage of total land used for agriculture	Percentage of total land used for agriculture	%	Permanent agriculture area from Eurostat 2011 - 2007 figures. CH 2005. Divided by total land area * 100
Percentage of economically active population employed full-time in agriculture	Percentage of economically active population employed full-time in agriculture	%	UNDP 2011 - extracted on Sep 19 from http://hdrstats.undp.org/en/tables/default.html
Percentage of population below the national poverty line	Percentage of population living below the national poverty line	%	UNDP 2011 - extracted on Sep 19 from http://hdrstats.undp.org/en/tables/default.html
Adult Literacy Rate	Percentage of adult literacy	%	UNDP 2011 - extracted on Sep 19 from http://hdrstats.undp.org/en/tables/default.html
Mean years of schooling	Adult mean years of schooling	years	UNDP 2011 - extracted on Sep 19 from http://hdrstats.undp.org/en/tables/default.html
Life Expectancy	Life expectancy at birth	years	UNDP 2011 - extracted on Sep 19 from http://hdrstats.undp.org/en/tables/default.html
Under 5 mortality	Under 5 mortality	persons/1000	UNDP 2011 - extracted on Sep 19 from http://hdrstats.undp.org/en/tables/default.html
Farms per square kilometer	Number of farms per square kilometer in the country	Farms/person	Total farms / Permanent agriculture area (FAO 2011)
Animals per farm	Number of units of each source per farm in the country.	Units/farm	Total animals of each type from FAOSTAT 2011 / Total farms from FAO 2011

Appendix I. Demonstration sheets provided along with the information sheets containing results of the cluster analyses

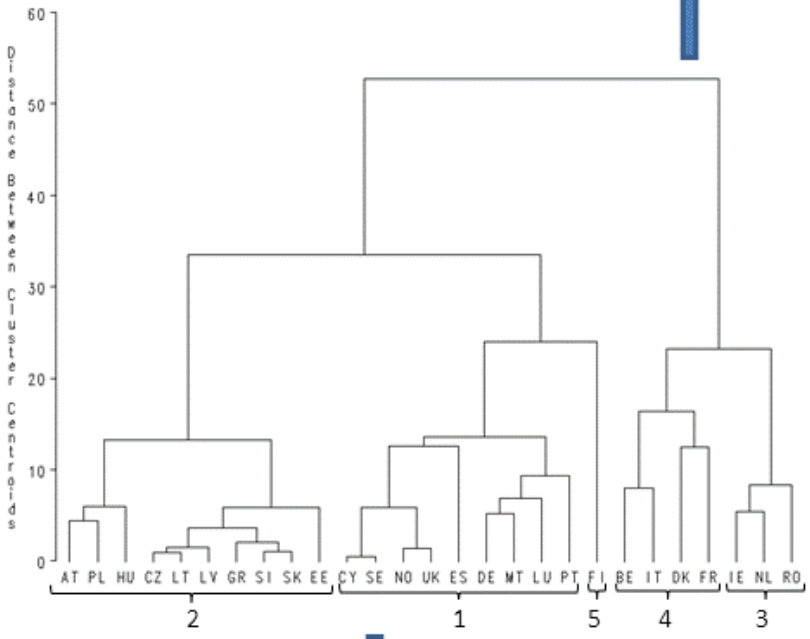
Demonstration sheet 1 - Components

Data table

Country	Enteritidis	Typhimurium	Others
ES	53.3	30.2	16.5
NO	58.9	22.7	18.4
UK	58.3	22.2	19.5
CY	61.9	17.7	20.3
SE	61.5	17.9	20.6
PT	69.6	26.1	4.3
LU	61.7	25.7	12.6
DE	64.0	30.4	5.6
MT	60.5	30.1	9.3
EE	86.3	9.2	4.5
GR	88.9	8.2	2.8
SI	90.2	6.3	3.5
SK	89.4	6.8	3.8
LV	91.8	5.0	3.2
CZ	93.1	4.9	2.0
LT	92.4	5.4	2.2
HU	76.2	13.6	10.2
AT	79.9	12.5	7.5
PL	82.3	9.1	8.7
RO	44.3	45.3	10.3
IE	42.4	38.8	18.8
NL	44.6	41.0	14.3
DK	30.4	52.6	17.0
FR	26.3	46.6	27.1
BE	24.5	65.1	10.4
IT	22.7	60.5	16.8
FI	49.0	17.6	33.5

Marks indicating values that drove the clustering of this group

Dendrogram



Line indicating cluster break

Illustration of the chosen cluster solution

Demonstration sheet 2 – what to look for in the table

Country	Enteritidis	Typhimurium	Others
ES	53.3	30.2	16.5
NO	58.9	22.7	18.4
UK	58.3	22.2	19.5
CY	61.9	17.7	20.3
SE	61.5	17.9	20.6
PT	69.6	26.1	4.3
LU	61.7	25.7	12.6
DE	64.0	30.4	5.6
MT	60.5	30.1	9.3
EE	86.3	9.2	4.5
GR	88.9	6.2	2.8
SI	90.2	6.3	3.5
SK	89.4	6.8	3.8
LV	91.8	5.0	3.2
CZ	93.1	4.9	2.0
LT	92.4	5.4	2.2
HU	76.2	13.6	10.2
AT	79.9	12.5	7.5
PL	82.3	9.1	8.7
RO	44.3	45.3	10.3
IE	42.4	38.8	18.8
NL	44.6	41.0	14.3
DK	30.4	52.6	17.0
FR	26.3	46.6	27.1
BE	24.5	65.1	10.4
IT	22.7	60.5	16.8
FI	49.0	17.6	33.5

Predominance of S.Enteritidis with a maximum difference of 45% to *S.* Typhimurium, and Typhimurium + Others still make up for 30-48%

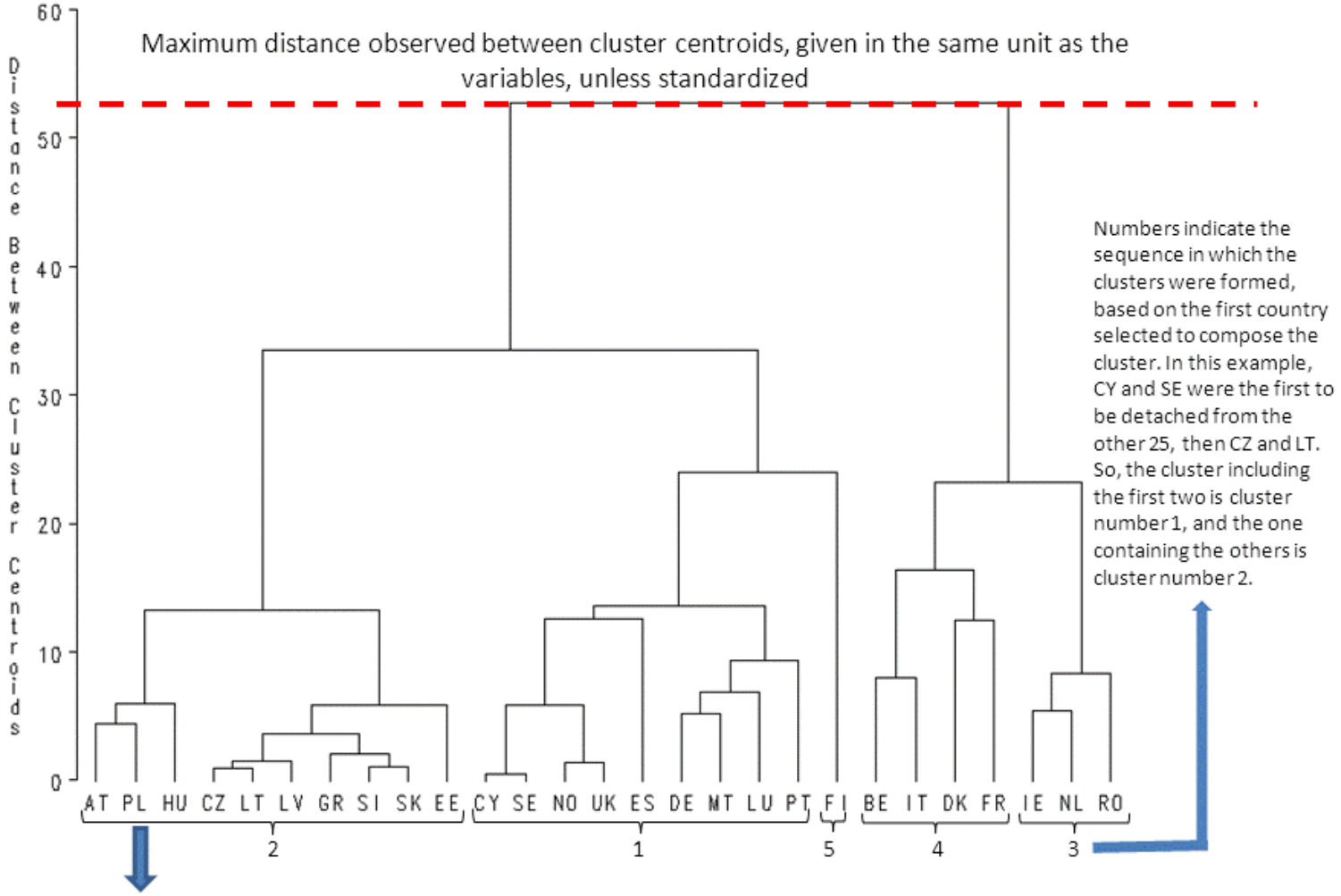
Absolute predominance of *S.* Enteritidis (>75%)

S.Enteritidis and *S.* Typhimurium roughly balanced

Predominance of *S.* Typhimurium

Predominance of *S.* Enteritidis with an expressive amount of "Others"

Demonstration sheet 3 – dendrogram

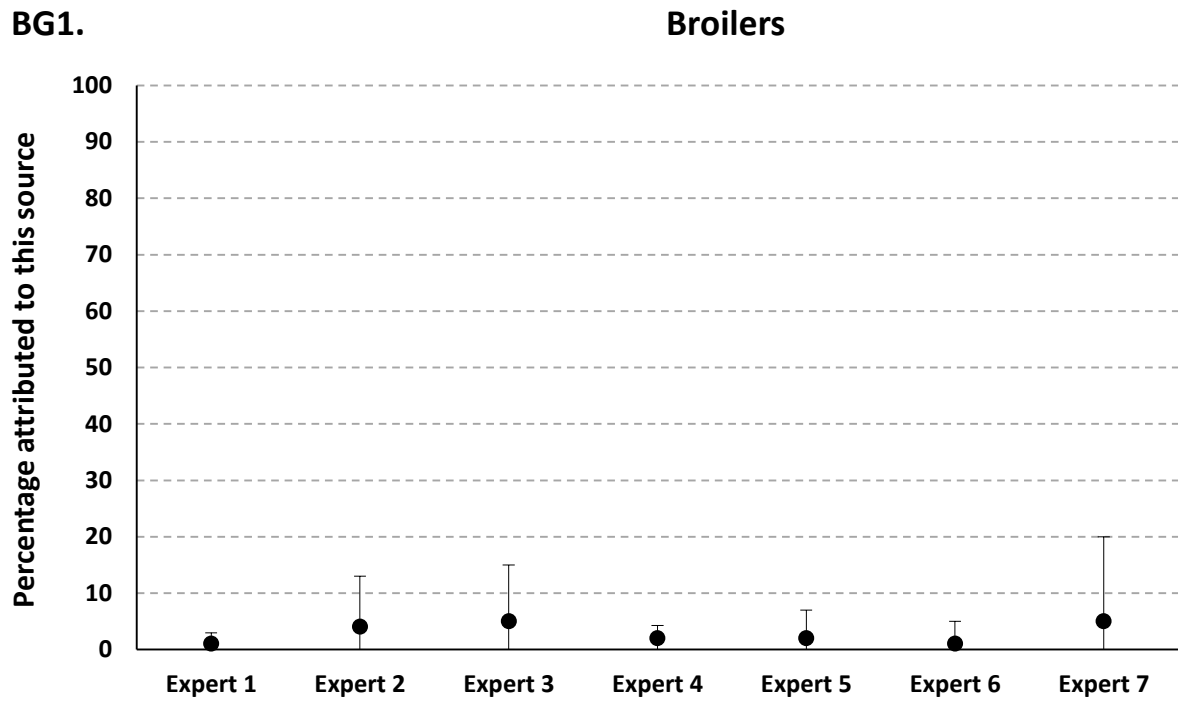


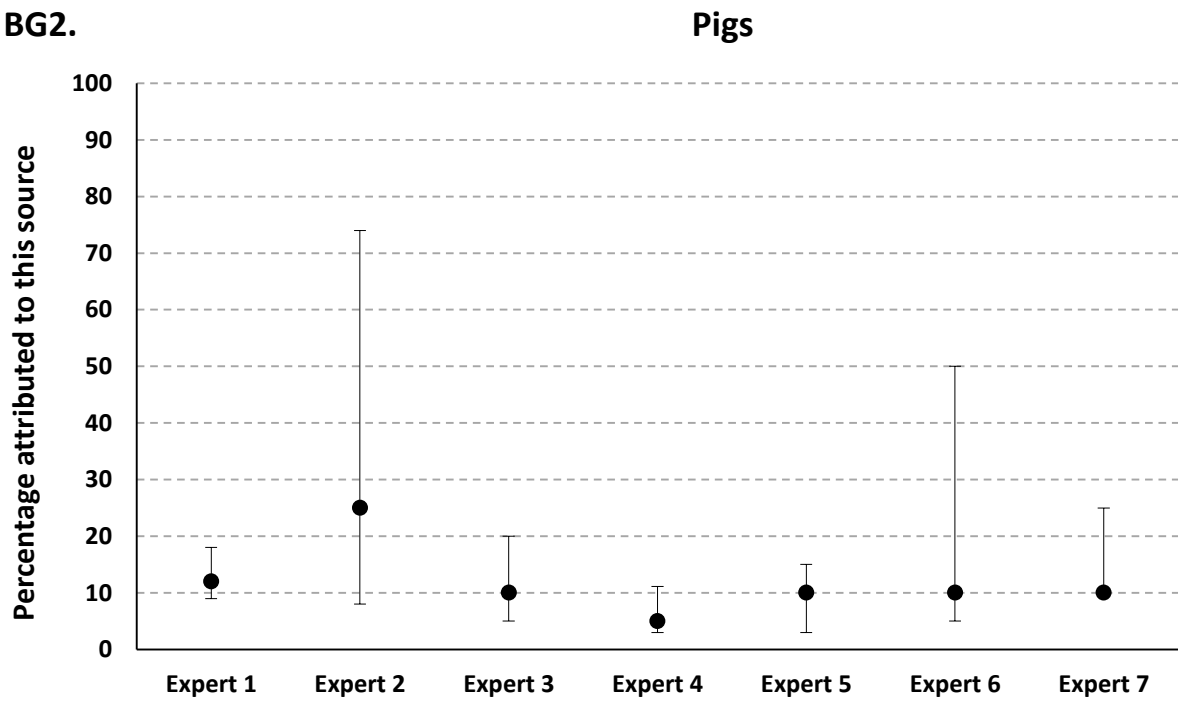
Although in the same cluster, AT and PL are more similar among themselves than to HU. The three of them are different from the others in their cluster, and CZ, LT, LV, GR, SI and SK are more similar among themselves than to EE.

Appendix J. Elicited estimates for Bulgaria, Norway and Romania

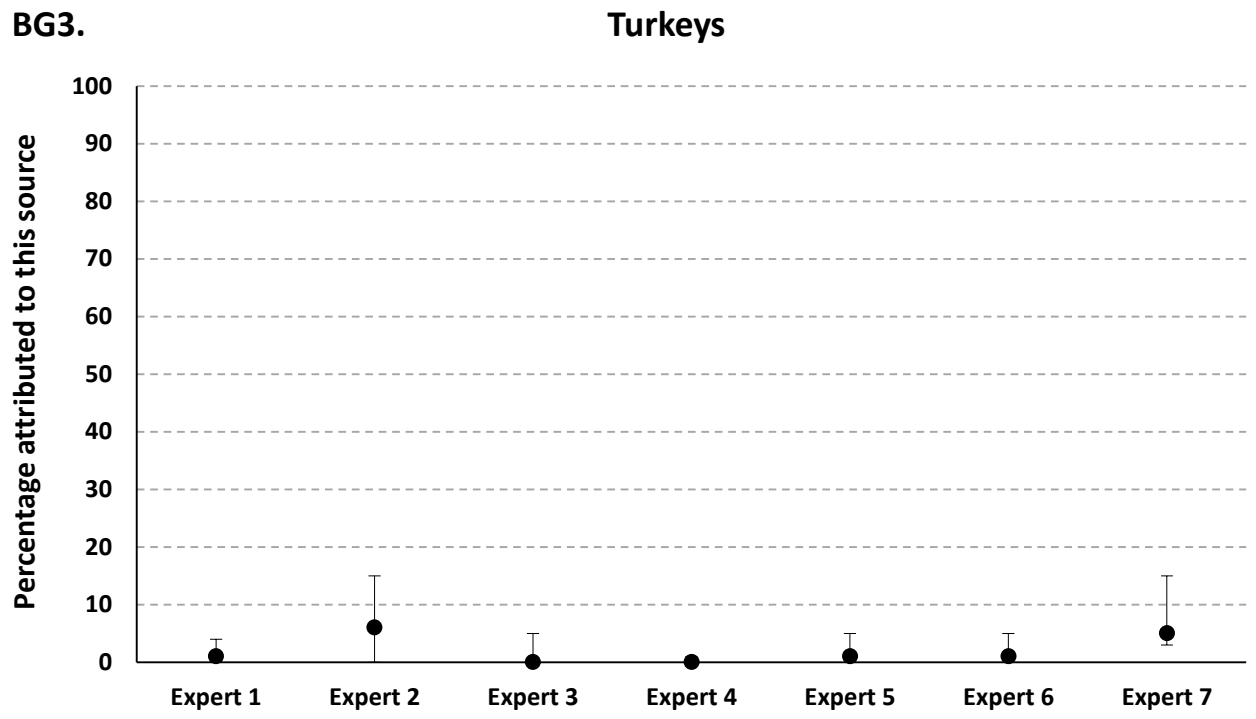
Respondant	Bulgaria			Norway			Romania		
Expert1	**Estimate**	**Min**	**Max**	**Estimate**	**Min**	**Max**	**Estimate**	**Min**	**Max**
Broilers	1	0.1	3	0.5	0.1	2	1	0.1	3
Pigs	12	9	18	5	1	10	12	9	18
Turkeys	1	0.1	4	1	0.1	2	1	0.1	4
Layers	70	5.5	85	2.5	0.1	5	70	5.5	85
Travel	2	0.1	5	80	70	90	2	0.1	5
Unknown/Other reservoirs	14	5	20	11	5	20	14	5	20
Expert2									
Broilers	4	0	13	5	0	13	4	0	13
Pigs	25	8	74	7	2	25	25	8	74
Turkeys	6	0	15	1	0	4	7	0	15
Layers	44	2	83	4	1	17	39	2	83
Travel	5	0	30	80	75	85	5	0	30
Unknown/Other reservoirs	17	4	38	4	2	11	20	4	38
Expert3									
Broilers	5	0	15	1	0	3	5	0	15
Pigs	10	5	20	4	1	10	15	5	40
Turkeys	0	0	5	1	0	3	5	0	15
Layers	75	60	90	4	1	10	70	30	80
Travel	0	0	5	80	70	90	0	0	15
Unknown/Other reservoirs	10	5	20	10	5	20	5	0	20
Expert4									
Broilers	2	0	4.3	1	0	1.9	2	0	4.3
Pigs	5	3	11.13	3	2	6.1	5	3	11.13
Turkeys	0	0	1	2	0.5	3.1	0	0	1
Layers	80	74	83	2	1.1	4.1	80	74	83
Travel	2	2	2	80	80	80	2	2	2
Unknown/Other reservoirs	11	4	12.5	12	3.1	15	11	4	12.5
Expert5									
Broilers	2	0	7	0.5	0	3	2	0	7
Pigs	10	3	15	5	0	10	25	17	30
Turkeys	1	0	5	1.5	0	5	1	0	5
Layers	80	70	90	2.5	0	7	50	40	65
Travel	2	0	7	80.5	75	90	1	0	3
Unknown/Other reservoirs	5	1	12	10	5	15	21	10	25
Expert6									
Broilers	1	0	5	1	0	2	1	0	15
Pigs	10	5	50	5	3	7	15	8	40
Turkeys	1	0	5	2	1	3	2	0	5
Layers	79	30	90	3	2	4	70	30	90
Travel	1	0	3	80	70	85	1	0	3
Unknown/Other reservoirs	7	3	10	9	3	20	11	5	15
Expert7									
Broilers	5	0	20	2	2	5	15	10	35
Pigs	10	10	25	10	10	15	-	-	-
Turkeys	5	3	15	35	5	50	-	-	-
Layers	52	40	80	8	3	10	-	-	-
Travel	8	5	10	25	15	40	-	-	-
Unknown/Other reservoirs	20	5	40	20	12	35	-	-	-

Appendix K. Individual expert guesses (most likely value, minimum and maximum) in Bulgaria (BG1 to BG4), Norway (NO1 to NO4) and Romania (RO1 to RO4).

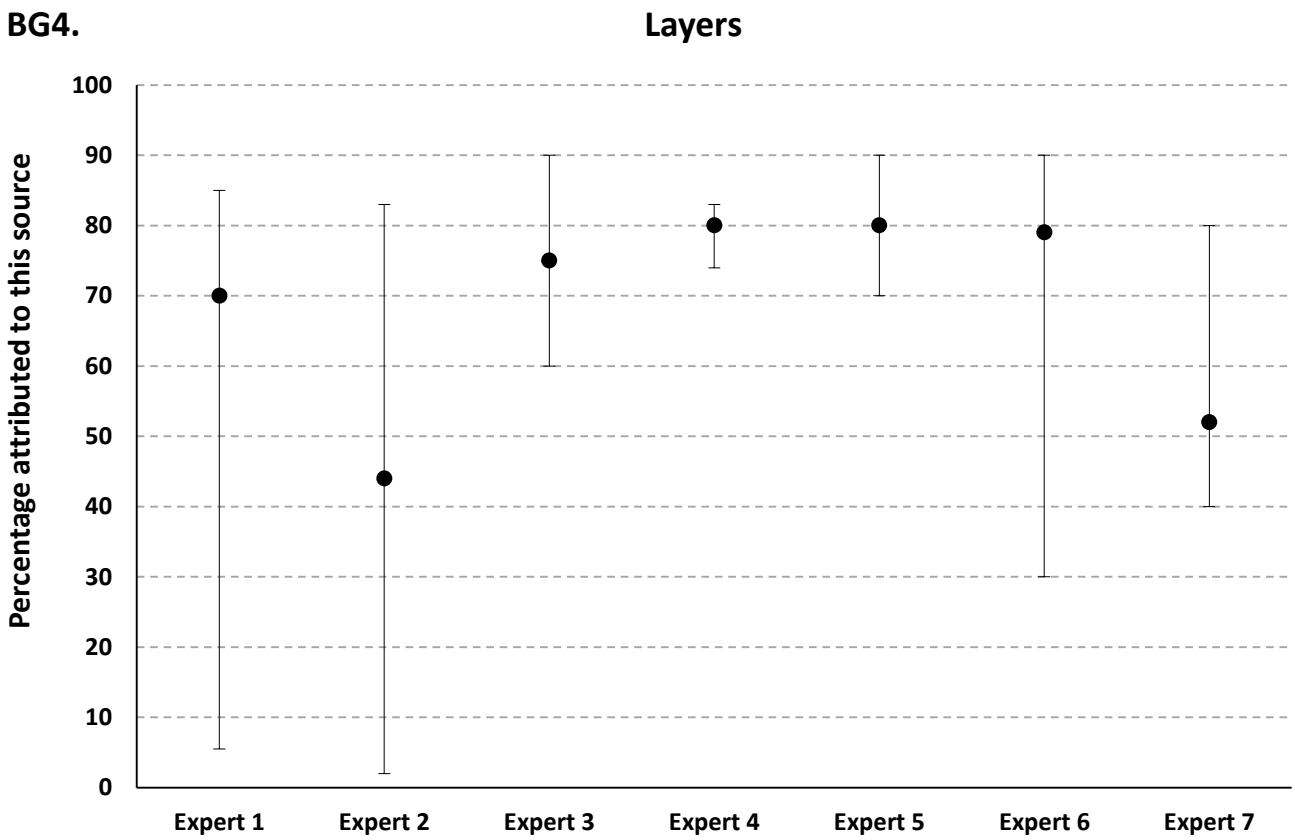




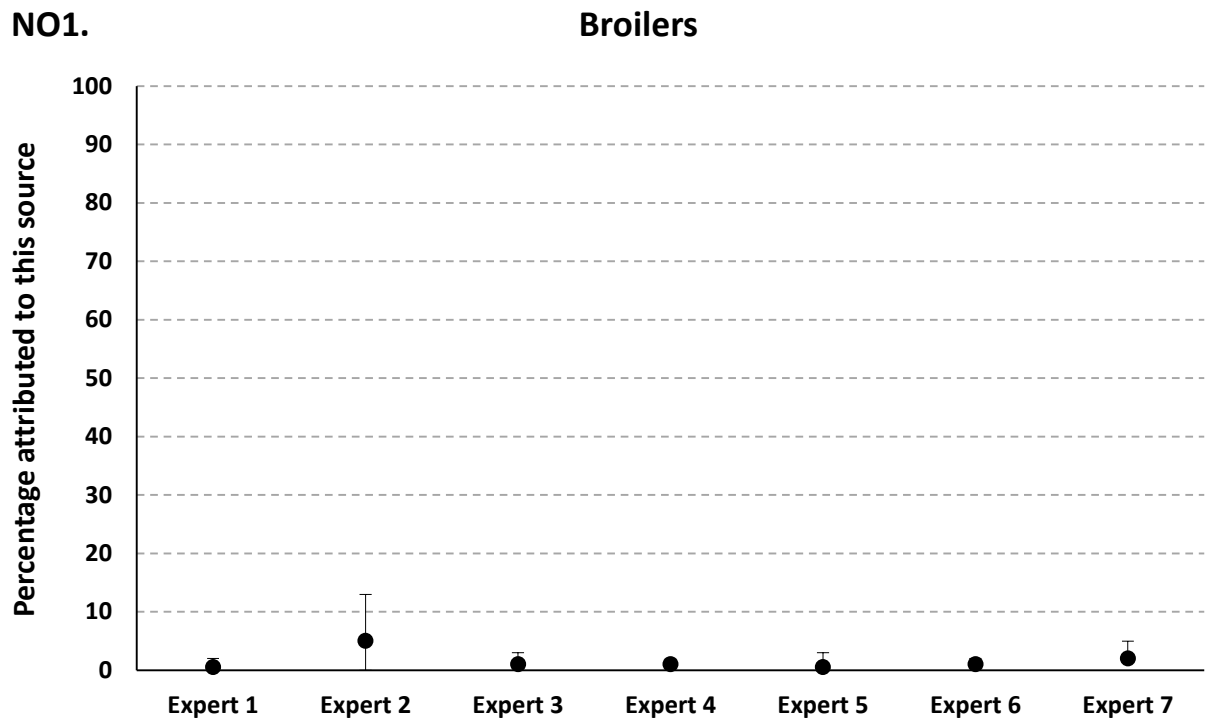
BG3.


Turkeys
Percentage attributed to this source
100
90
80
70
60
50
40
30
20
10
0
Expert 1
Expert 2
Expert 3
Expert 4
Expert 5
Expert 6
Expert 7

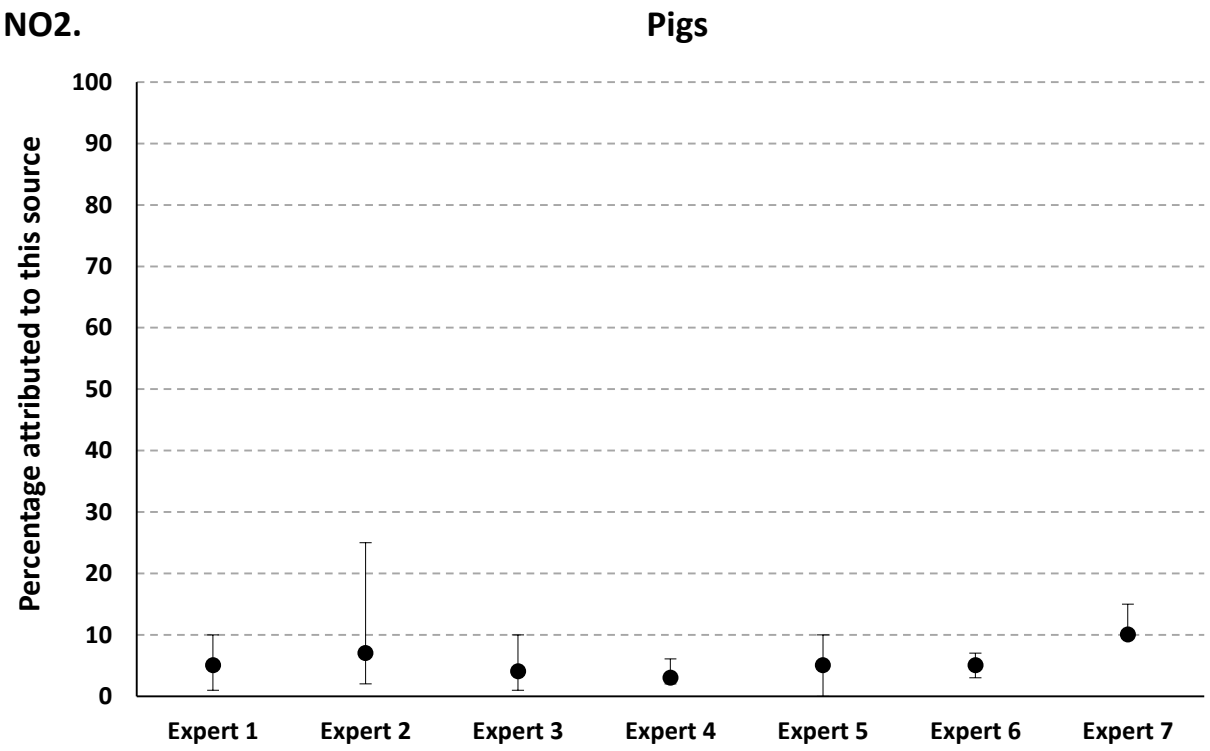
BG4.


Layers
Percentage attributed to this source
100
90
80
70
60
50
40
30
20
10
0
Expert 1
Expert 2
Expert 3
Expert 4
Expert 5
Expert 6
Expert 7

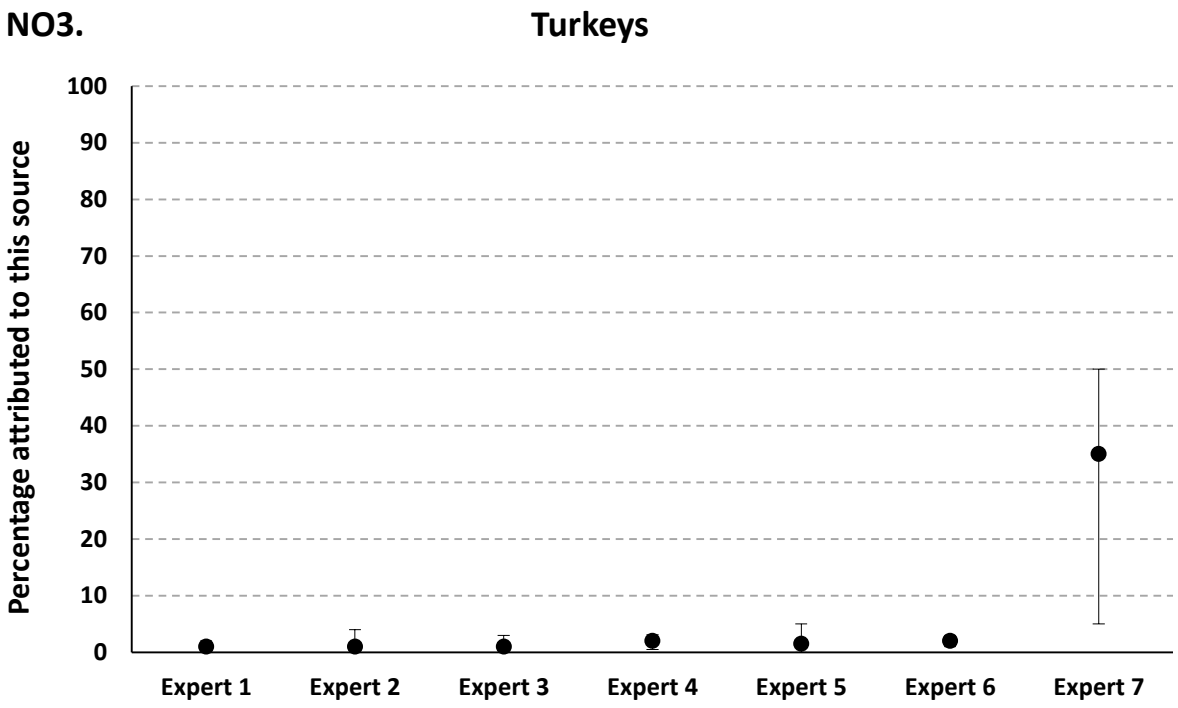
NO1.


Broilers
Percentage attributed to this source
100
90
80
70
60
50
40
30
20
10
0
Expert 1
Expert 2
Expert 3
Expert 4
Expert 5
Expert 6
Expert 7

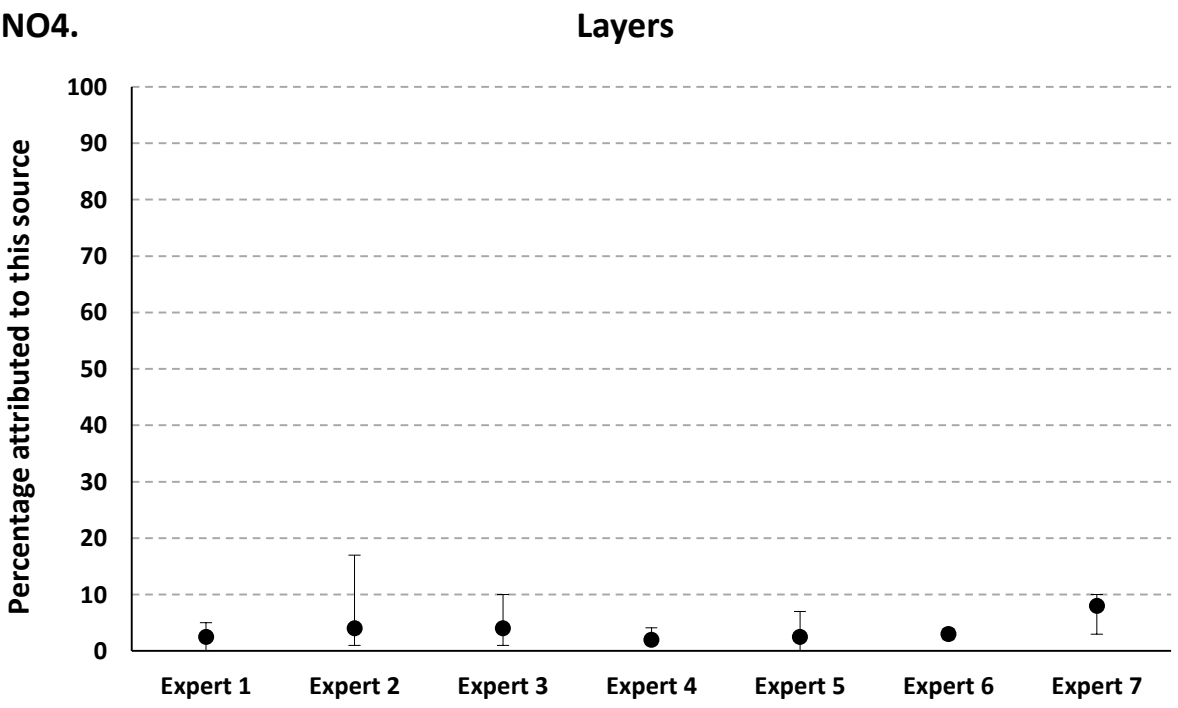
NO2.


Pigs
Percentage attributed to this source
100
90
80
70
60
50
40
30
20
10
0
Expert 1
Expert 2
Expert 3
Expert 4
Expert 5
Expert 6
Expert 7

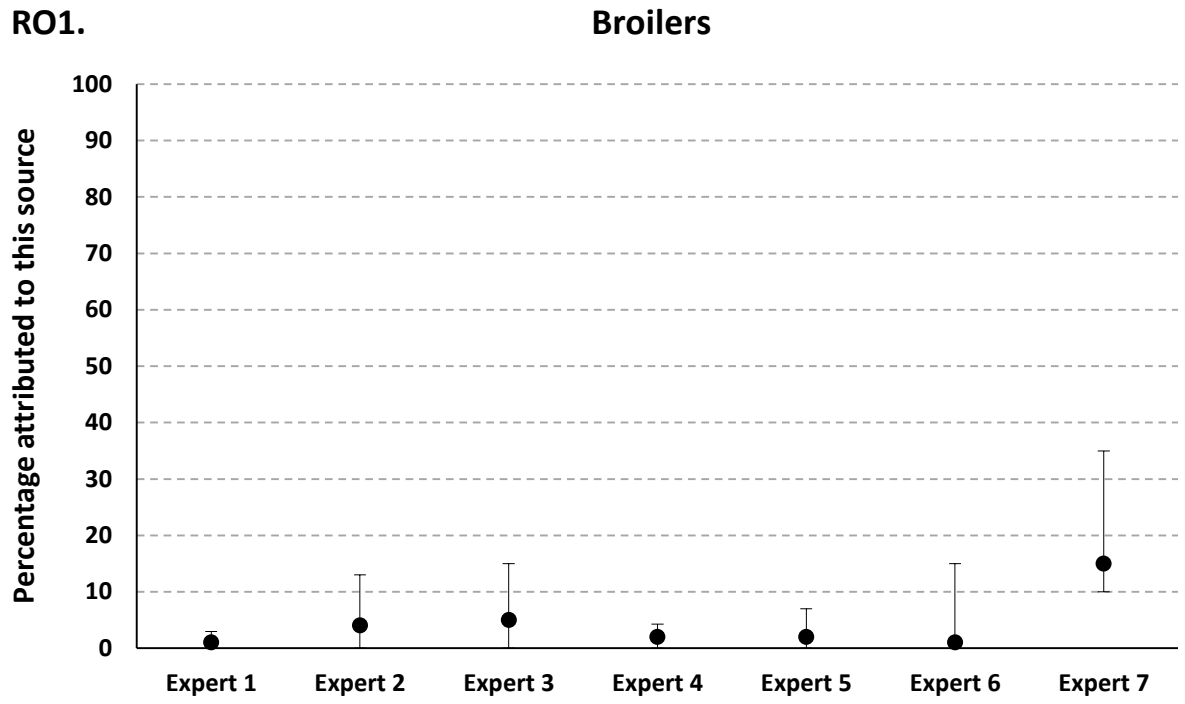
NO3.


Turkeys
Percentage attributed to this source
100
90
80
70
60
50
40
30
20
10
0
Expert 1
Expert 2
Expert 3
Expert 4
Expert 5
Expert 6
Expert 7

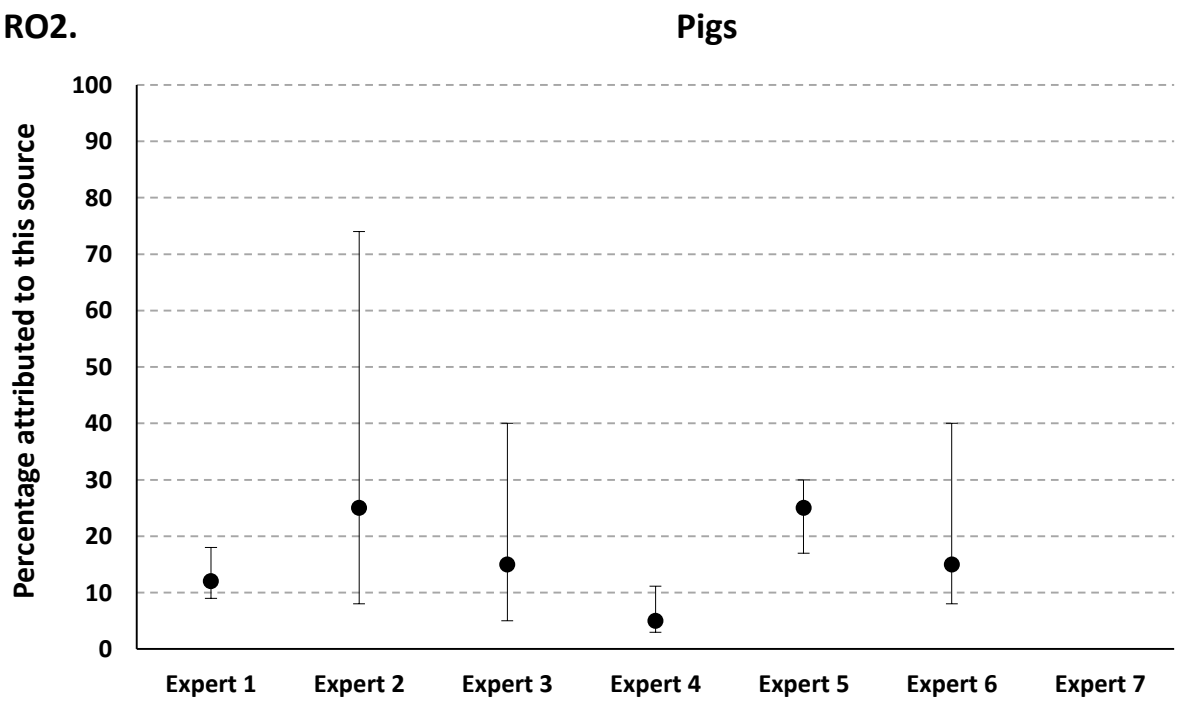
NO4.


Layers
Percentage attributed to this source
100
90
80
70
60
50
40
30
20
10
0
Expert 1
Expert 2
Expert 3
Expert 4
Expert 5
Expert 6
Expert 7

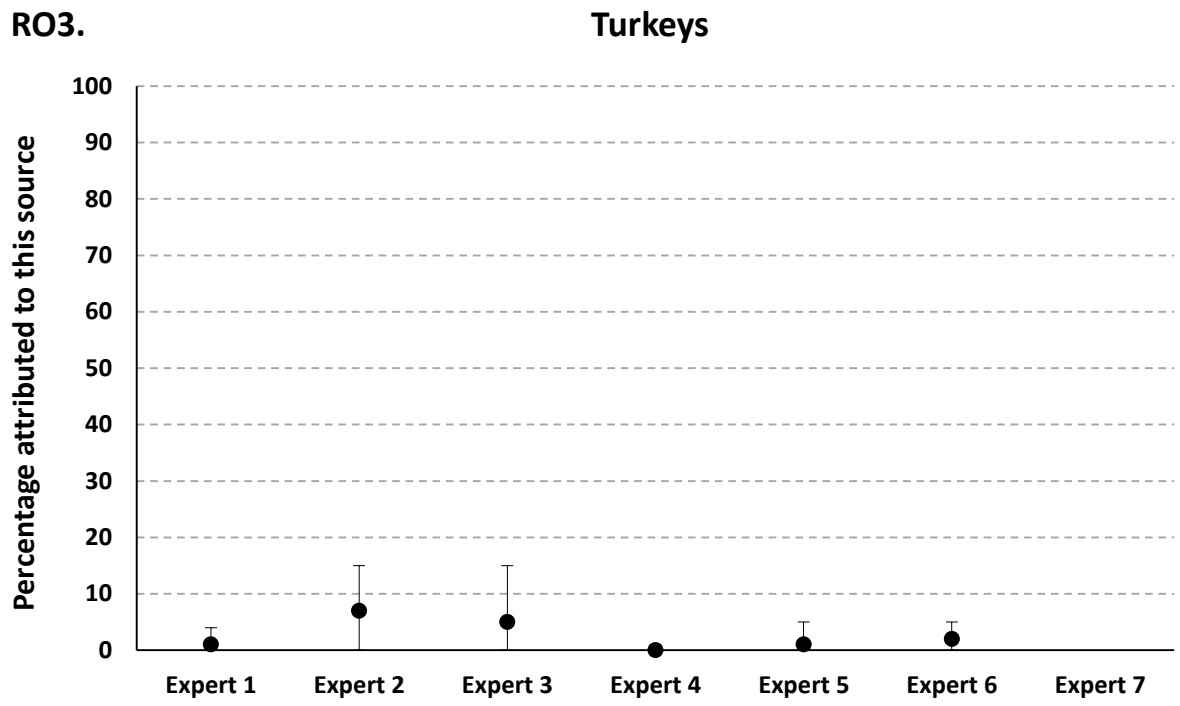
RO1.


Broilers
Percentage attributed to this source
100
90
80
70
60
50
40
30
20
10
0
Expert 1
Expert 2
Expert 3
Expert 4
Expert 5
Expert 6
Expert 7

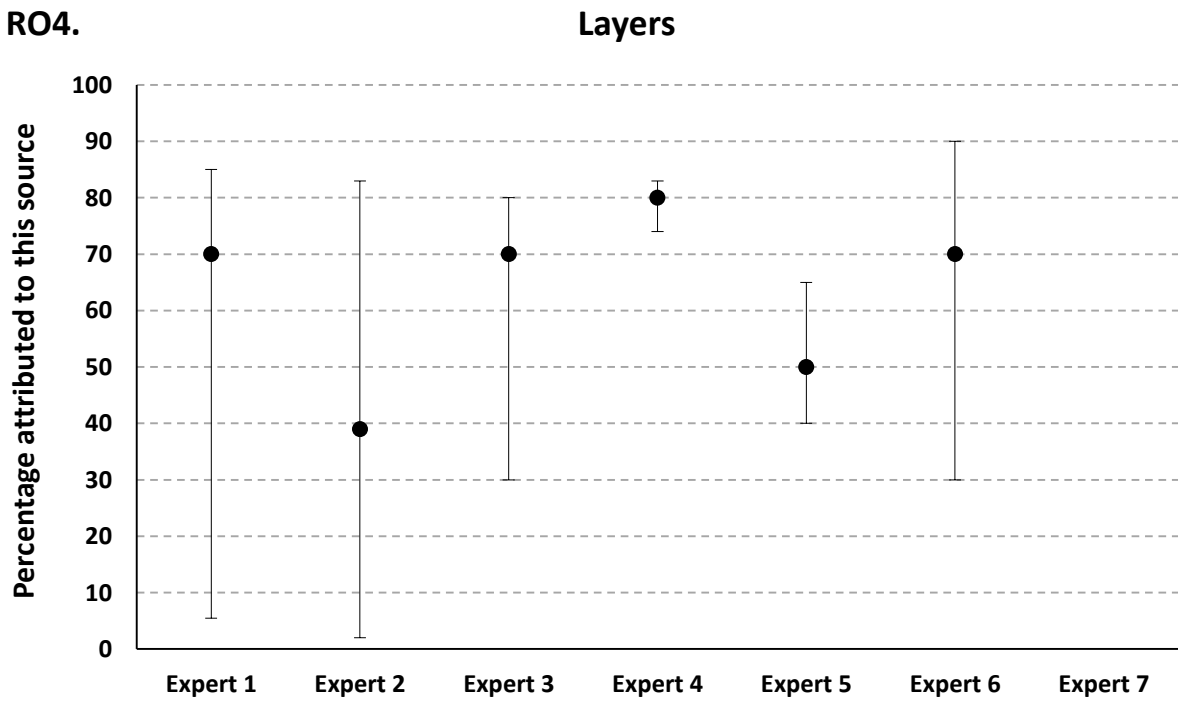
RO2.


Pigs
Percentage attributed to this source
100
90
80
70
60
50
40
30
20
10
0
Expert 1
Expert 2
Expert 3
Expert 4
Expert 5
Expert 6
Expert 7

RO3.

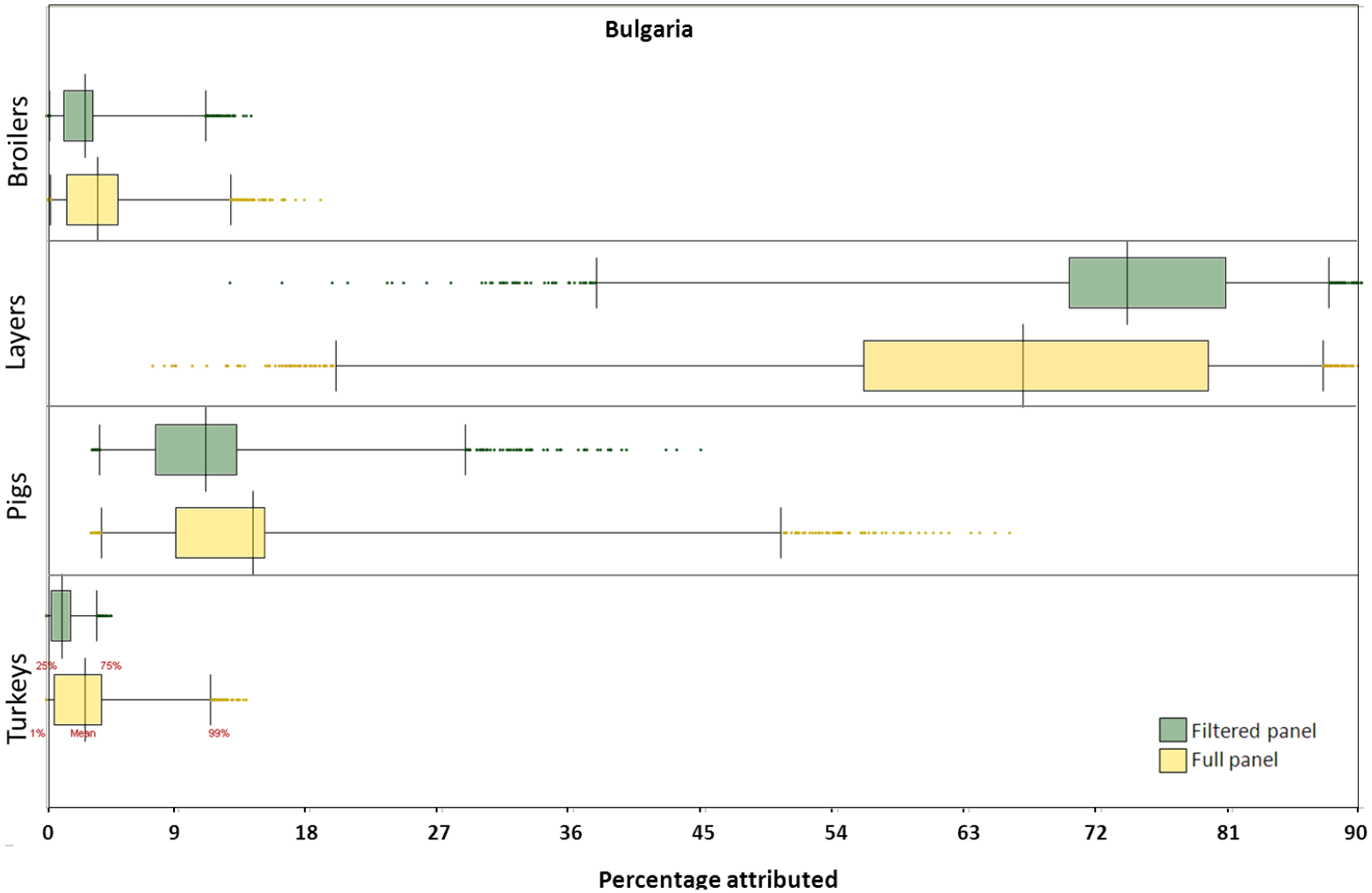

Turkeys
Percentage attributed to this source
100
90
80
70
60
50
40
30
20
10
0
Expert 1
Expert 2
Expert 3
Expert 4
Expert 5
Expert 6
Expert 7

RO4.

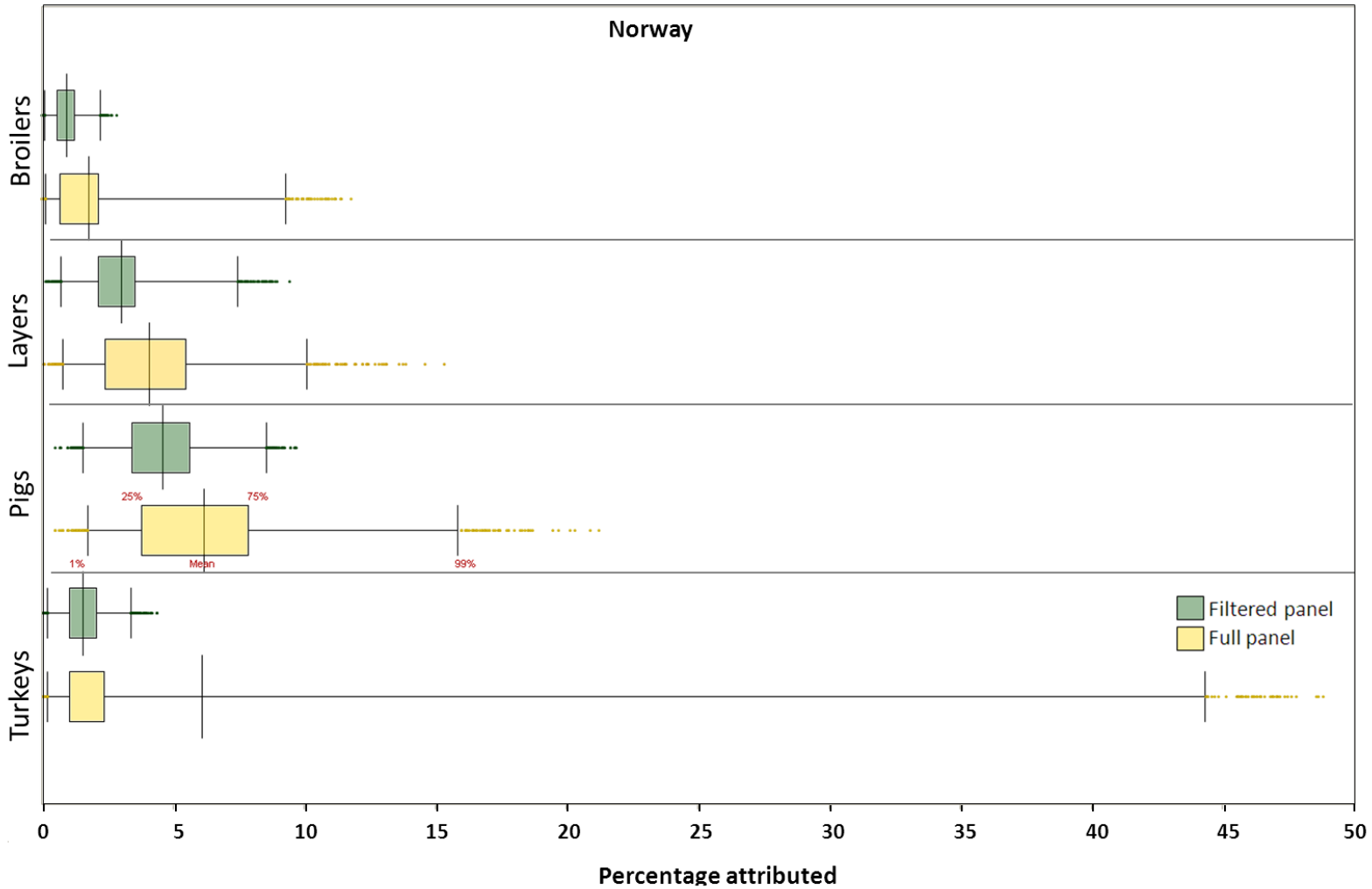

Layers
Percentage attributed to this source
100
90
80
70
60
50
40
30
20
10
0
Expert 1
Expert 2
Expert 3
Expert 4
Expert 5
Expert 6
Expert 7

Appendix L. Joint panel estimate distributions in Bulgaria (L1), Norway (L2) and Romania (L3)

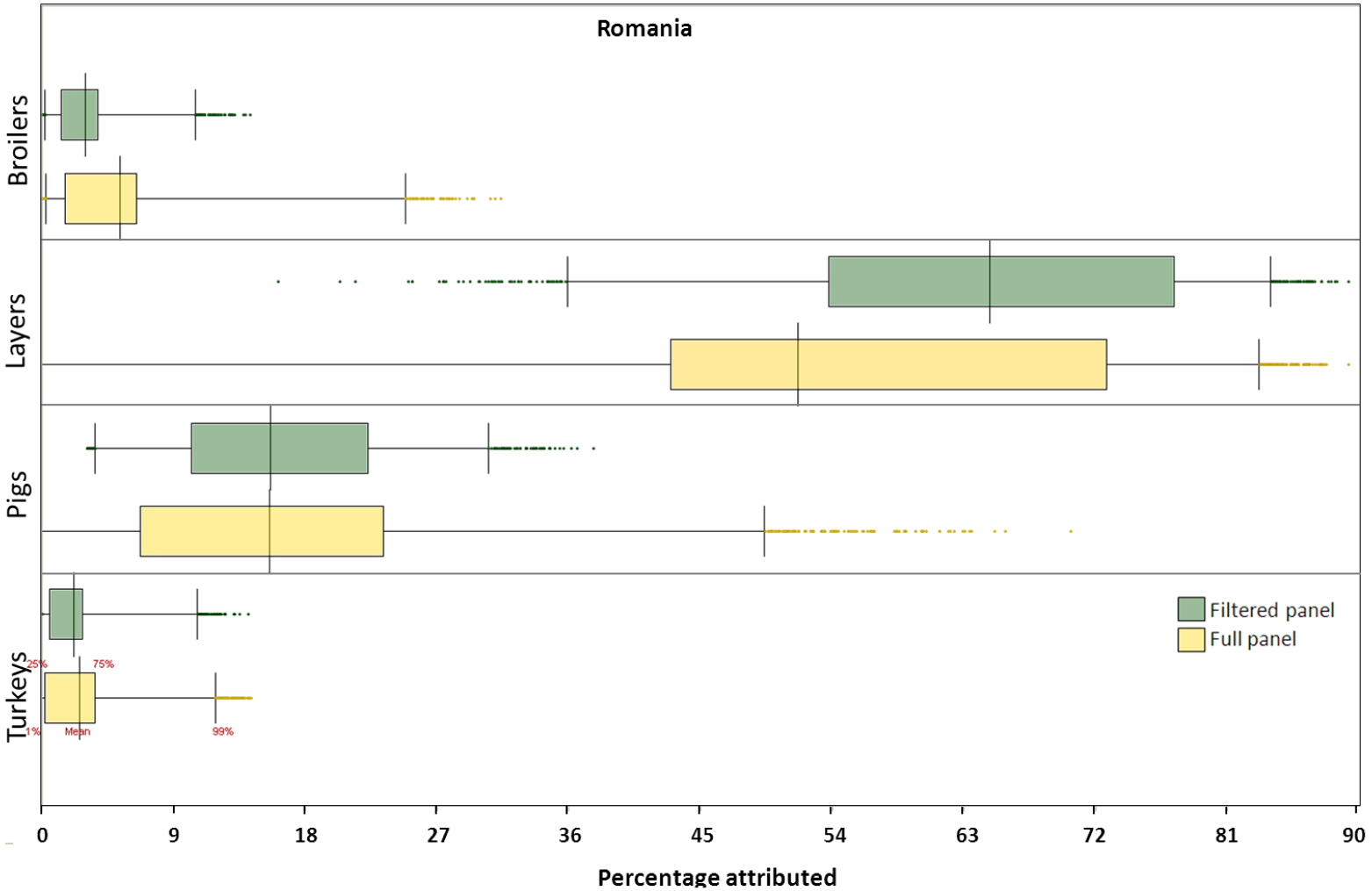
L1. Joint estimate distributions from the full and filtered panels in Bulgaria.



L2. Joint estimate distributions from the full and filtered panels in Norway



L3. Joint estimate distributions from the full and filtered panels in Romania



National Food Institute
Technical University of Denmark
Mørkhøj Bygade 19
DK - 2860 Søborg

Tel. 35 88 70 00
Fax 35 88 70 01

www.food.dtu.dk

ISBN: 978-87-92763-46-4